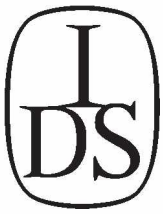


INSTITUT FÜR DEUTSCHE SPRACHE

Forschungsinfrastrukturen in außeruniversitären Forschungseinrichtungen

Forschungsbericht

Norman Fiedler
Antonina Werthmann
Maik Stührenberg
Oliver Schonefeld
Joachim Bingel
Andreas Witt

Forschungsinfrastrukturen in außeruniversitären Forschungseinrichtungen

1. Digitale Forschungsinfrastrukturen

1.1 Einleitung

Forschungsprimärdaten, die aus Messungen, Umfragen oder Erhebungen entstehen können, bilden seit jeher die unverzichtbare Grundlage jeder empirisch arbeitenden Forschungsrichtung. Seit dem Erscheinen digitaler Technologien und Medien und immer stärkerer elektronischer Vernetzung ist das Aufkommen solcher Daten unterschiedlichster Beschaffenheit in Gestalt digitaler Ressourcen exponentiell gestiegen. Selbst vor den sich von den Methoden und Ansätzen technischer, lebens- und naturwissenschaftlicher Disziplinen bisweilen fundamental unterscheidenden und oft als konservativ geltenden Geistes- und Sozialwissenschaften hat diese Entwicklung nicht halt gemacht. Digitale Technologien in der Datenverarbeitung und -erzeugung haben auch in diesen Bereichen Einzug gehalten und das Selbstverständnis der „Humanities“ hin zu einer sehr viel stärker empirisch arbeitenden Forschung nachhaltig beeinflusst (Henrich 2011).

Dieses deutlichen Wandels unbenommen lässt sich ein Gesamtbild der Anwendung digitaler Methoden aufgrund starker disziplinärer und mithin methodischer Zersplitterung und, gerade mit Blick auf Datenformate und softwaregestützte Methoden, noch immer sehr ausgeprägter Heterogenität in einem Bereich der Forschung nur sehr schwer zeichnen, der vorwiegend durch die Individualität seiner Vertreter geprägt zu sein scheint.

Das Beispiel der germanistischen Sprachwissenschaften, für die stellvertretend die jüngsten Aktivitäten des Instituts für Deutsche Sprache (IDS) im Bereich digitaler Sprachressourcen und der elektronischen Erschließung der deutschen Sprache mit Methoden der Computerlinguistik herausgegriffen werden, soll an dieser Stelle stellvertretend für ähnliche Institutionen stehen. Daher ist das Ziel dieses Dokuments die Darstellung des aktuellen Stands am Beispiel des IDS. Die hier gemachten Aussagen lassen sich daher nicht unverändert auf andere Einrichtungen übertragen, können aber als Richtlinien und Empfehlungen für ähnliche außeruniversitären Forschungseinrichtungen[1] mit geistes- oder sozialwissenschaftlichem Schwerpunkt in Hinblick auf die Forschungsinfrastruktur dienen.

[1] Unter außeruniversitären Forschungseinrichtungen, die in Deutschland in den Forschungsgemeinschaften Fraunhofer, Helmholtz, Max-Planck und Leibniz zusammengefasst sind, werden Institutionen verstanden, die im Unterschied zu Universitäten in der Zahl ihrer Mitarbeiter oft kleiner sind, sich andererseits aber auf bestimmte Forschungsbereiche konzentrieren, oft wissenschaftliche Serviceleistungen anbieten und sich nur in geringem Umfang der Lehre widmen.

1.2 Anforderungen an eine Forschungsinfrastruktur

Unter dem Begriff *Forschungsinfrastrukturen* wird gemeinhin die Gesamtheit der Daten, Informationen und Ressourcen für die Wissenschaft und das Forschungsmanagement sowie alle an ihre Bereitstellung, Pflege und Aufbewahrung gekoppelten Dienstleistungen und Forschungen zusammengefasst. Dazu zählen neben Großgeräten auch Wissensressourcen wie digitale und analoge Sammlungen, Archive und Datenbanken (Wissenschaftsrat 2011a).

> „Als Forschungsinfrastrukturen werden diejenigen teilweise einzigartigen ‚Einrichtungen, Ressourcen und Dienstleistungen' in öffentlicher oder privater Trägerschaft verstanden, die speziell für wissenschaftliche Zwecke errichtet, mittelfristig bis tendenziell permanent bereitgestellt werden und für deren sachgerechte Errichtung, Betrieb und Nutzung in der Regel spezifische fachwissenschaftliche oder interdisziplinäre (Methoden-)Kompetenzen erforderlich sind. Ihre Funktion ist es, Forschung, Lehre und Nachwuchsförderung zu ermöglichen oder zu erleichtern. Sie sind örtlich fixiert, auf mehrere Standorte verteilt oder werden ohne definierte physische Anlaufstelle ausschließlich virtuell bereitgestellt. Sie werden nicht ausschließlich von einzelnen Personen oder Gruppen genutzt, sondern stehen prinzipiell einer internationalen Fachgemeinschaft oder mehreren Fachgemeinschaften offen." (Wissenschaftsrat 2011b)

In einer Zeit fortschreitender Digitalisierung sind Aufbau und Pflege von Forschungsinfrastrukturen eine so zentrale Aufgabe einer Forschungsinstitution, dass die Umsetzung dieser Maxime eine wesentliche Antwort auf künftige Herausforderungen der Wissenschaft sein wird.

Je nach wissenschaftlichem Schwerpunkt einer Forschungseinrichtung werden auch die Anforderungen an Forschungsinfrastrukturen unterschiedlich gestellt. Sie hängen nicht zuletzt davon ab, ob die Seite der Anbieter oder die der Nutzer betrachtet wird.

Anforderungen eines Nutzers/Wissenschaftlers an Forschungsinfrastrukturen**:**

- Wissenschaftlicher Mehrwert,
- transparenter und benutzerfreundlicher Datenzugang,
- offene Architektur und Flexibilität für disziplinäre Spezifika,
- Unterstützung und Beratung in technischen Fragen,
- Sicherheit, Nachhaltigkeit und Reproduzierbarkeit von Forschungsdaten,
- gemeinsame Erzeugung, Analyse und Bearbeitung von Forschungsdaten,
- Kompatibilität der Infrastrukturen und Schnittstellen,
- Umsetzungen der Maßgaben für gute wissenschaftliche Praxis,
- Rechtssicherheit, Technische (Ausfall-)Sicherheit von Systemen und Daten.

Die Anforderungen eines Anbieters an Forschungsinfrastrukturen**:**

- Klares wissenschaftliches Anforderungsprofil,
- klare Zuständigkeiten zwischen Anbieter und Nutzer sowie transparente Entscheidungsprozesse,
- klar dokumentierte und strukturierte Datenproduktion seitens der Wissenschaftler,

- eindeutige Nutzungsbestimmungen,
- Einhaltung von Standards,
- Einhaltung von Fristen und Regelungen.

Forschungsinfrastrukturen sollten unter Berücksichtigung der Ressourcen sämtlicher im weitesten Sinne datenhaltender Organisationseinheiten eines Forschungsinstituts wie Bibliothek, IT-Abteilung, Archiv(e), Verwaltung sowie von einzelnen Fachabteilungen bzw. Projekten durch Koordinierung und Zusammenarbeit aller Beteiligten aufgebaut werden, sodass sich innerhalb aller datenverarbeitenden Einrichtungen eines Forschungsinstituts unnötige Doppelstrukturen in einer abgestimmten und einheitlichen Organisationsstruktur vermeiden lassen.

Jeder dieser Bereiche wird im Folgenden eingehender analysiert und beschrieben. Es steht dabei außer Frage, dass es zahlreiche Schnittmengen zwischen den einzelnen Bereichen gibt, die nicht isoliert betrachtet werden können. Jede Forschungsinfrastruktur sollte jedoch organisatorisch so aufgestellt sein, dass unnötige Überschneidungen vermieden und einer Desorganisation entgegen gewirkt wird.

1.3 Kategorien von Forschungsinfrastrukturen

In seinen Empfehlungen zu Forschungsinfrastrukturen in den Geistes- und Sozialwissenschaften weist der Wissenschaftsrat (WR) auf die Bedeutung von gut aufgestellten und adaptiven Forschungsinfrastrukturen als wesentliche Voraussetzung für eine leistungsstarke und international wettbewerbsfähige Forschung hin (Bundesministerium für Bildung und Forschung 2013). Im Bereich der Geistes- und Sozialwissenschaften ist insbesondere die Arbeit mit Forschungsdatensammlungen sowie ihre Digitalisierung, Archivierung und langzeitige Aufbewahrung ein vorrangiges Anliegen, infolgedessen der Bedarf nach entsprechenden digitalen Lösungen perspektivisch größer zu werden verspricht.

Der Wissenschaftsrat unterscheidet hierbei grundsätzlich vier Forschungsinfrastrukturkategorien:

- **Großgeräte:** Forschungsschiffe, -fluggeräte, -satelliten, etc.;
- **Informationstechnische und e-Infrastrukturen:** Grid- oder Cloud-Netzwerke, Rechenzentren, Kooperationsverbünde für die Hard- und Softwarenutzung sowie entsprechenden Support;
- **Soziale Infrastrukturen:** Im Zentrum einer sozialen Forschungsinfrastruktur stehen *Begegnungsräume*, die für den persönlichen kommunikativen Austausch unter Wissenschaftlern und Wissenschaftlerinnen und die Entwicklung von neuen oder aktuellen Fragestellungen zur Verfügung gestellt werden, z.B. Tagungszentren oder Forschungslabore (Hohoff 2011; Wissenschaftsrat 2011a). Soziale Forschungsinfrastrukturen fördern gemeinschaftliche Arbeit *ad personam*, bilden kommunikative Netzwerke zwischen Wissenschaftlern, begünstigen den Dialog zwischen Forschern und dem wissenschaftlichen Nachwuchs und dienen der „Forschungskommunikation über alle Fragen eines Faches“ (Wissenschaftsrat 2011a, S. 68). Diese Ebene unterliegt jedoch anderen Kategorien und wird in diesem Dokument nicht näher betrachtet. Infrastrukturen wie Bibliotheken und Museen werden in dieser Studie nicht un-

ter ihrem sozialen Aspekt der Wissensgenerierung durch (experimentelle) Interaktion betrachtet, sondern vielmehr als Plattformen der Datenhaltung und Dienstleistung ähnlich den Forschungsinstituten oder als institutionalisierte Forschungsinfrastrukturen verstanden.

- **Informationsinfrastrukturen:** Der Begriff der Informationsinfrastrukturen umfasst ähnlich wie die zuvor genannten sozialen Infrastrukturen die für Forschung und Lehre im Sinne einer Grundversorgung relevanten Primärdaten, Wissensdatenbanken, wissenschaftliche Sammlungen, Netzstrukturen, Archive oder Bibliotheken.

Trotz des zweifelsfreien technischen Fortschritts in den Geistes- und Sozialwissenschaften stellt der Einsatz von Großgerätschaften wie Schiffen oder Satelliten noch immer eine Seltenheit dar. Zumeist werden solch kostenintensive Infrastrukturen dann auch nicht selbst angeschafft und betrieben, sondern auf vorhandene Ressourcen zurückgegriffen. Soziale Infrastrukturen fallen zwar weitgehend in den Ereignishorizont dieser Studie, doch ist die spezielle Perspektive dieser Kategorie mit dem Fokus auf Wissenserzeugung vor dem Hintergrund der hier fokussierten Fragestellung weniger relevant. So stehen im Folgenden vor allem die Kategorien informationstechnische Infrastrukturen (Rechnerinfrastruktur) und Informationsinfrastrukturen (Informationsversorgungsysteme) im Mittelpunkt der Betrachtungen und sollen auf ihre Wechselbeziehung zueinander untersucht werden.

2. Informationstechnische Infrastruktur

Unter Berücksichtigung dediziert wissenschaftlicher Anforderungen gibt es keine klare und eindeutige Definition von informationstechnischen Infrastrukturen. Grundsätzlich subsumiert dieser sehr weit gefasste Begriff sämtliche Maßnahmen und Rahmenbedingungen für den Aufbau und die Organisation von rechnergestützter Forschung sowie Übertragung und Verarbeitung von für die Forschung relevanten Informationen. Hierzu zählen u.a. auch Aufgabenfelder wie Telefonie, Netzwerke, Gebäudetechnik usw. Diese vielfältige Anwendbarkeit des Begriffs der IT-Infrastruktur führt dazu, dass terminologische wie thematische Überschneidungen zwischen den hier angesprochenen Kategorien von Forschungsinfrastrukturen kaum zu vermeiden sind.

Im Folgenden werden die wesentlichen Bereiche der IT-Infrastruktur näher beschrieben, die für den Aufbau im Rahmen einer außeruniversitären Forschungseinrichtung eine Rolle spielen:

- Langfristige IT-Strategie
- Hardware und Software
- Kosten
- Verfügbarkeit & Datensicherung („Backup“)
- Arbeitsplätze

2.1 Entwicklung einer langfristigen IT-Strategie

In Zeiten massiven Vordringens digitaler Methoden in alle Forschungsfelder hängen in jeder Forschungseinrichtung die Aktualität, Zugänglichkeit, Wirkung, Nachvollziehbarkeit und Nachhaltigkeit ihrer Ergebnisse zumindest mittelbar von einer effizienten und stabilen IT-

Infrastruktur ab. Zur Erreichung dieses Ziels ist eine umfassende IT-Strategie ein erster und bedeutender Schritt. Die Entwicklung und Umsetzung einer solchen IT-Strategie stellt sich äußerst komplex und aufwendig dar und reicht von Konzepten zur Sicherheit der Daten bis hin zur Ausfallsicherheit des gesamten Systems. Um den zukünftigen Herausforderungen im IT-Bereich zu begegnen, sollte an Leitbild, Kompetenzen, Leistungsangebot (Prozesse, Services), Leistungserbringung, IT-Architektur und Strategieumsetzung schon frühzeitig gearbeitet werden.

Eine IT-Strategie definiert somit den strategischen Leitfaden für das tägliche Handeln in allen Tätigkeitsfeldern und Kernkompetenzen der Einrichtung. Die Kernelemente einer solchen Strategie umfassen insbesondere die folgenden Punkte:

- Aufgabendefinition,
- Services und Leistungsspektrum,
- Nutzungsbestimmungen und -einschränkungen,
- Zuständigkeiten, Aufbau und Finanzierung,
- Technische Ausstattung,
- Umsetzung von Standards und Empfehlungen,
- Datensammlung und -erstellung,
- Datenzugang, -speicherung und -bearbeitung,
- Datennutzung und -interpretation,
- Langfristige Archivierung und Sicherung,
- Support.

Als Grundlage für die Entwicklung einer IT-Strategie können die Empfehlungen der Kommission für IT-Infrastruktur der DFG dienen (DFG 2010).

2.2 Identitäts-Management (IdM)

Ein wichtiger Aspekt der IT-Infrastruktur ist ein Identitäts-Management-Systems (IdM), das es ermöglicht die Benutzeridentitäten innerhalb der IT-Infrastruktur der Einrichtung global und transparent zu verwalten. Die Verwendung eines IdM erhöht die Sicherheit der Forschungsrichtung, erleichtert den Administratoren bei der Benutzerverwaltung die Arbeit und hilft damit Kosten zu senken (Gietz 2004). Die Anwendung eines IdM wird vom DFN-Verein daher auch als ein grundlegender Bestandteil der Infrastruktur einer Einrichtung gesehen und ist als eine Voraussetzung zur Teilnahme an der DFN-AAI[2] (DFN-AAI 2010b) unumgänglich. Daher sollte ein IdM ein integraler Bestandteil der IT-Strategie sein.

2.3 Hardware und Software

Unter dem Begriff Hardware wird die gesamte Rechentechnik (z.B. PCs, Laptops, Storage-Systeme etc.), die Netzwerktechnik (z.B. Router, Switches, Kabel), Peripheriegeräte (Tastatur, Bildschirm, Drucker, Scanner etc.) sowie weitere Geräte, die zu deren Betrieb benötigt werden (z.B. Server- bzw. Netzwerkschränke, unterbrechungsfreie Stromversorgungen, etc.) zusammengefasst. Es lässt sich jüngst beobachten, dass die Anforderungen an die Ausstat-

[2] Im Abschnitt 2.10.2 wird auf den Themenkomplex AAI eingegangen.

tung von Forschungsinstituten mit Hardwarekomponenten stetig zunehmen. Eine große Rolle spielen hierbei Aspekte der Funktionalität sowie der höheren Performanz, Rechenleistung, längeren Lebensdauer und nicht zuletzt der Kompatibilität mit anderen Hardwarekomponenten, Daten oder Programmen.

Unter Software wird die Vielfalt von Programmen zusammengefasst, die die Hardware erst für den Endanwender sinnvoll nutzbar machen. Hier wird vor allem unterschieden zwischen Systemsoftware, d.h. vorrangig Betriebssystemen, ohne die grundsätzlich kein Rechner nutzbar ist, und Anwendungssoftware, d.h. Programmen, die den Benutzerinnen und Benutzern bestimmte Funktionalitäten bereitstellen. Letzterer Bereich lässt sich noch feiner untergliedern in Standardsoftware, wie z.B. Office-Pakete, und Spezialsoftware, die sehr spezifisch auf die Anwendungsdomäne zugeschnitten bzw. oft sogar speziell dafür entwickelt worden ist.

2.4 IT-Kosten

Aufbau und Betrieb einer IT-Infrastruktur sind mit hohen Kosten verbunden. Grundsätzlich setzen sich die Kosten aus den Aufwendungen der Anschaffung und des laufenden Betriebs zusammen. Unter den Anschaffungskosten lassen sich die Mittel zusammenfassen, die für den Einkauf von Hardware und Software benötigt werden, um eine IT-Infrastruktur aufzubauen.

Die laufenden Kosten fallen regelmäßig über die Dauer des Betriebs an und lassen sich in Arbeits- und Wartungskosten unterteilen. Während Arbeitskosten die Gehälter der Administratoren sowie IT-Mitarbeiter umschreiben, benennen Wartungskosten sämtliche Instandhaltungsarbeiten an der Infrastruktur, aber auch mögliche Gebühren für die Aufrechterhaltung des Betriebs, wie z.B. Software-Updates, Reparatur, Wartung, Lizenzen etc. Nicht zu vernachlässigen sind darüber hinaus die für den Betrieb der IT-Hardware und ihrer Klimatisierung benötigten Energiekosten. Der Energieverbrauch kann je nach Stand der zugrundeliegenden Technik und dem Stellenwert von Green-IT in der IT-Strategie eines Instituts neben Personalkosten einen nicht zu unterschätzenden Anteil an den Gesamtbetriebskosten einnehmen und wird angesichts steigender Strompreise perspektivisch eher zunehmen als stagnieren. Im Bereich der IT-Infrastruktur sind aus diesem Grund zusehends innovative Konzepte und Lösungen gefragt, die zur effizienten Nutzung von Energie führen. Einsparungspotenzial bietet beispielsweise die Berücksichtigung des Kriteriums Energieeffizienz im Entscheidungsprozess für die Anschaffung von Hardware. Auch Cloud-Computing, also die Bereitstellung verschiedener IT-Leistungen wie Software oder Plattformen für Entwicklung und Betrieb von Anwendungen auf Basis virtueller Umgebungen, kann zur Senkung des Energiebedarfs und der Gesamtkosten beitragen (vgl. Abschnitt 2.5.2).

Zur Optimierung der Kosten empfiehlt sich zunächst die Etablierung eines transparenten Abrechnungssystems in Gestalt von Einzelposten für Personalaufwand, Betriebs- und Verbrauchskosten des IT-Bereichs, um vor dem Hintergrund gründlicherer Evaluationen, Rechnungsprüfungen und Vollkostenrechnungen tatsächliche Bedarfe genauer abschätzen zu können. Im Anschluss daran existieren zahlreiche individuell umsetzbare Strategien. So trägt die konsequente Nutzung von freier oder Open-Source-Software zwar nachhaltig dazu bei Anschaffungs- und Lizenzkosten niedrig zu halten. Es ist allerdings auch zu bedenken, dass

diese zum einen nur einen vergleichsweise geringen Betrag ausmachen und eventuelle Umschulungskosten der Mitarbeiter der Ersparnis gegenüber stehen können. Ein nicht zu unterschätzender Vorteil im Gegensatz zu kommerziellen Produkten ist allerdings die Möglichkeit, quelloffene Software selbst an die eigenen Anforderungen anpassen zu können.[3] Kommerzielle Software kann auch an die eigenen Bedürfnisse angepasst werden, allerdings ist dies meist mit zusätzlichen Kosten durch den Hersteller bzw. einen Dienstleister verbunden.

Die IT-Kosten in einer Forschungseinrichtung können weiterhin durch Kooperation mit anderen Einrichtungen (gemeinsame Beschaffungen, Auslagerung, Rahmenverträge) und die Mitgliedschaft in Verbünden (s. Abschnitt 2.10) zur gemeinsamen Nutzung von Daten, Software und anderen Ressourcen gesenkt werden.

2.5 Netzwerke und Sicherheit

Das Internet bildet das weltweit größte aus einzelnen, selbstständigen Computernetzwerken bestehende Konstrukt, das den nahezu freien Datenaustausch ermöglicht. Daten und Informationen stehen hier prinzipiell jedem Nutzer offen, der über die notwendigen technischen Voraussetzungen verfügt – damit ist es zu einem unverzichtbaren Arbeitsmittel für Forschungsinstitutionen geworden (Lux 2005).

Der Zugang zum Internet erfolgt über ein Wide Area Network (WAN) mit Hilfe eines Internet Service Providers (ISP), bei universitären Forschungseinrichtungen üblicherweise durch eine Anbindung ans Deutsche Forschungsnetz (DFN).[4]

Ein Intranet (LAN – Local Area Network) ist im Gegensatz zum Internet nicht öffentlich, sondern wird innerhalb von Institutionen zur internen Vernetzung eingesetzt, um den Austausch von Informationen und die Kommunikation untereinander zu ermöglichen. Es bleibt in seiner Zugänglichkeit auf einen festgelegten Nutzerkreis beschränkt. In außeruniversitären Forschungseinrichtungen können Nutzer über das Intranet beispielsweise auf Dateiserver, betriebsinterne öffentliche Informationen wie Betriebsvereinbarungen, Mitarbeiterzeitschriften, Vordrucke, Vorlagen, Formulare oder Dokumente zugreifen. Darüber hinaus können über das Intranet weitere zentrale Anwendungsangebote wie z.B. Terminkoordination, gemeinsame Dokumentenbearbeitung usw. bereitgestellt werden. Art und Umfang, in dem Informationen oder Serviceangebote über ein Intranet zur Verfügung gestellt werden, werden kaum über technische oder gar rechtliche Schranken reguliert und obliegen damit der Verantwortung jedes einzelnen Forschungsinstituts.

Ein LAN ist meist entweder durch eine klassische Verkabelung per Ethernet oder per WLAN realisiert. Oft werden beide Techniken auch kombiniert, d.h. reguläre Arbeitsplätze per Ethernet und Notebooks per WLAN angebunden. Einzelne Arbeitsplätze lassen sich in Ausnahmefällen auch über alternative Verbindungen wie Powerline oder andere Funkübertragungsverfahren (wie z.B. Wireless Personal Area Networks oder WiMAX) anbinden. Allerdings werden, besonders bei den Funkverfahren, nicht die Sendeleistungen, Reichweite und

[3] Auch hierbei sollte bedacht werden, dass Anpassungen nicht nur einmalig erfolgen, sondern oftmals bei Aktualisierung der zu Grunde liegenden Software erneut durchgeführt werden müssen. Hierzu ist entsprechendes Knowhow innerhalb der Forschungseinrichtung unverzichtbar.
[4] https://www.dfn.de/

Datenübertragungsrate eines sorgfältig eingerichteten WLAN erreicht (Bundesamt für Sicherheit in der Informationstechnik 2009).

Die Reichweite eines WLAN-Accesspoints ist abhängig von seiner Umgebung. Die Wände des Gebäudes, Bäume, Elektrogeräte und Anzahl der Benutzer können sie entscheidend beeinflussen. Außerdem gilt, dass die Signalstärke (und damit die Übertragungsrate) mit der Entfernung zum Accesspoint abnimmt. Die Funkzelle kann allerdings durch mehrere sich überlappende Accesspoints erweitert werden.

In Bezug auf die Organisation bringt ein Funknetzwerk viele Vorteile mit sich: Zum einen ist keine Neuverkabelung notwendig (mit Ausnahme der Anbindung an das WAN) und ein bestehendes Netzwerk kann kostengünstig erweitert werden. Das ist besonders dann vorteilhaft, wenn eine Verkabelung zwischen Gebäuden oder Gebäudeteilen benötigt wird, z.B. falls eine IT-Infrastruktur auf mehrere Standorte verteilt ist. Auch externe Nutzer können die bestehende Netzwerkstruktur ohne großen Aufwand nutzen. Überdies wird die Mobilität und Flexibilität eigener Mitarbeiter verbessert, die nicht an einem bestimmten Arbeitsplatz und an einem PC gebunden sind. Zu beachten ist allerdings, dass ein WLAN bei Weitem nicht so robust gegenüber äußeren Störungseinflüssen ist wie ein klassisches LAN. Aus diesem Grund sollte bei der Planung von Funknetzen entsprechende Back-up- und Notfallkonzepte für sicherheitskritische Forschungsinfrastrukturen (z.B. zentrale Server) vorgesehen werden.

Bezüglich der Sicherheit der IT-Infrastruktur einer Institution sind vorrangig zwei Punkte zu beachten: Einerseits ist ein nicht gewollter Zugriff auf Systeme, Prozesse oder Daten zu verhindern. Andererseits unterliegt die Integrität von Daten einer sicheren und kontinuierlichen Funktionalität der Systeme, die durch menschliches Versagen oder technischen Defekt kompromittiert werden kann. Der erste Schritt hin zu einer sicheren IT-Infrastruktur ist daher eine umfassende Risikoanalyse zur Abschätzung der Gefahrenwahrscheinlichkeit für die eigenen Systeme und Gerätschaften sowie eine Risikofolgenabschätzung zum Grad der zu erwartenden Schäden und Einschränkungen der Arbeitsabläufe des Instituts. Aus den Ergebnissen dieser Vorabeinschätzung sind eine Hierarchisierung der Sicherungswürdigkeit von Daten und Systemen sowie ein angemessenes Vorgehen abzuleiten.

Die Sicherheitsmaßnahmen werden in der Regel in den Sicherheitsrichtlinien (oft auch als „Security Policy“ bezeichnet) der IT-Sicherheitsverantwortlichen und Datenschutzbeauftragten definiert und dokumentiert und sollten für alle Mitarbeiter verbindlich sein (ISO/IEC JTC 1/SC 27 2005; Bundesamt für Sicherheit in der Informationstechnik 2008a).

Die Kernpunkte einer IT- Sicherheitsstrategie sind im Folgenden aufgeführt:

- Definition der Sicherungswürdigkeit von Daten (s. a. Abschnitt 5.2),
- Erkennung und Verhinderung von Gefährdungen durch Schadsoftware (Viren, Trojaner) und Angriffen auf die Netzwerkinfrastruktur; Ausstattung aller Arbeitsplätze mit Schutzsoftware (Antivirus- und Antispamprogramme, Firewall-Systeme) und deren Pflege,
- Betriebsvereinbarungen zum Umgang mit Forschungsdaten und zur Nutzung der Infrastruktur zu privaten Zwecken (s. Abschnitt 2.7),

- Vermeidung von Datenverlusten durch technische Pannen oder personelle Fehler („Backup-Strategie“, s. Abschnitt 2.6),
- Strategien für den Zugriff auf und die Auffindbarkeit von Forschungsdaten,
- Netzwerksicherheit (inkl. WLAN) und Zugriffssicherheit über externe Zugänge (s. Abschnitt 2.5.3),
- Verschlüsselung von Daten.

Gerade in Funknetzwerken sind Verschlüsselungsverfahren notwendig, um die Übertragung der Daten und Informationen zu sichern, Maßnahmen wie MAC-Adressenfilter (Media-Access-Control-Filter) können allenfalls als zusätzliche Vorkehrungen getroffen werden.

Verschlüsselungsverfahren stellen einerseits sicher, dass nicht autorisierten Dritten der Zugang zum drahtlosen Kommunikationssystem verwehrt wird und andererseits die ausgesendeten Informationen unverfälscht den Empfänger erreichen. Es muss grundsätzlich vermieden werden, in einer IT-Infrastruktur unverschlüsselte, kabellose Netze aufzubauen, die es Unbefugte ermöglichen unbemerkt Angriffe auf IT-Infrastruktur einer Forschungseinrichtung durchzuführen.

Um Angriffe zu vermeiden oder zu erschweren, wurden Sicherheitsmaßnahmen auch innerhalb des WLAN-Standards IEEE 802.11 entwickelt, z.B. Robust-Security-Network (RSN) oder WPA/WPA2. Mit fortschreitender Entwicklung der Technik ist jedoch damit zu rechnen, dass weitere Sicherheitsmaßnahmen entwickelt und implementiert werden müssen, um Angriffe auf Informationen und Daten zu erschweren. Da sich allerdings die potentiellen Angriffsvektoren ebenfalls ändern, kann eine absolute Sicherheit nie gewährleistet werden.

2.5.1 Sicherheit am Arbeitsplatz

Die Mitarbeiter der Forschungseinrichtung sollten für das Thema IT-Sicherheit sensibilisiert werden. Dies kann zum Beispiel durch Schulungen oder Informationsveranstaltungen erreicht werden, auch können forschungseinrichtungsinterne Leitfäden mit Maßnahmen zur Unterstützung der IT-Sicherheit an die Mitarbeiter ausgegeben werden. Ein solcher Leitfaden kann beispielsweise folgende Punkte thematisieren:

- **Aktueller Softwarestand:** Sofern dies nicht bereits durch technische Maßnahmen der IT-Abteilung gewährleistet ist, sollten alle Mitarbeiter dafür Sorge tragen, dass auf ihren Computerarbeitsplätzen ein aktuelles Betriebssystem eingesetzt wird, alle Sicherheits-Updates regelmäßig eingespielt werden und ein Virenscanner mit regelmäßig aktualisierten Virensignaturen eingesetzt wird. Weiterhin ist darauf zu achten, dass besonders auch Programme wie Web-Browser (inkl. Plug-Ins wie Flashplayer, Java, etc.), Mail-Client, PDF-Betrachter und Office-Programme auf einem aktuellen Stand sind.
- **Mail-Anhänge:** Auch wenn alle eingehenden E-Mails üblicherweise auf dem Posteingangsserver der Institution nach Viren gescannt werden, sollten Mitarbeiter beim Öffnen von Anhängen – insbesondere von unbekannten Kontakten – Vorsicht walten lassen.

- **Arbeitsplätze sperren:** Die Mitarbeiter sollten ihre PCs beim Verlassen des Arbeitsplatzes sperren. Diese Maßnahme verhindert den unbefugten Zugriff auf den Rechner (und in Folge auf Forschungsdaten).
- **Passwortsicherheit:** Mitarbeiter sollten dazu angehalten werden, ihre Passwörter regelmäßig zu ändern. Es sollte selbstverständlich sein, dass Passwörter nicht aufgeschrieben werden oder auf sonstiger Weise Dritten zugänglich gemacht werden. Durch ein Identity-Management-System (IdM), besonders in Kombination mit Single-Sign-On (SSO), d.h. bei einmaliger Passworteingabe während des Anmeldeprozesses, kann ein guter Kompromiss zwischen Bequemlichkeit und Sicherheit erreicht werden.
- **Private Geräte (BYOD – Bring Your Own Device):** Private Geräte wie Notebooks, Smartphones etc. sollten nur dann im Institutsnetzwerk verwendet werden, wenn dies explizit erlaubt und durch eine entsprechende Richtlinie durch die Institution geregelt ist. Sind solche Geräte zugelassen, sollten diese dann einem Sicherheitsstandard entsprechen, der mit den regulären Arbeitsplätzen vergleichbar ist, d.h. es sollte Wert darauf gelegt werden, dass aktuelle Betriebssysteme und Virenscanner verwendet werden, dass Sicherheits-Updates eingespielt sind, etc. Der Zugang zum lokalen Netzwerk (z.B. per WLAN) sollte vorzugsweise über ein separates Netzwerk erfolgen. Solch ein separates Netzwerk erleichtert es, Einschränkungen in der Nutzung von Instituts-Ressourcen festzulegen und technisch durchzusetzen, z.B. über die obligatorische Nutzung von Terminalservern.

2.5.2 Sicherheit beim Cloud-Computing

„Cloud Computing bezeichnet das dynamisch an den Bedarf angepasste Anbieten, Nutzen und Abrechnen von IT-Dienstleistungen über ein Netz. Angebot und Nutzung dieser Dienstleistungen erfolgen dabei ausschließlich über definierte technische Schnittstellen und Protokolle. Die Spannbreite der im Rahmen von Cloud Computing angebotenen Dienstleistungen umfasst das komplette Spektrum der Informationstechnik und beinhaltet unter anderem Infrastruktur (z.B. Rechenleistung, Speicherplatz), Plattformen und Software.“ (Bundesamt für Sicherheit in der Informationstechnik 2012, S. 15).

Grundsätzlich kann eine Cloud drei verschiede Arten von Diensten anbieten:

- **Infrastructure as a Service (IaaS):** Rechenleistung, Datenspeicher oder Netze werden als Dienst angeboten. Dabei verwaltet der Nutzer diese Infrastruktur selbst.
- **Platform as a Service (PaaS):** Dem Nutzer wird neben der kompletten Infrastruktur auch ihre Verwaltung bereitgestellt. Auf dieser Plattform können mit standardisierten Schnittstellen des Anbieters eigene Programme bereitgestellt und benutzt werden.
- **Software as a Service (SaaS):** Hier wird durch den Cloud-Anbieter eine bestimmte Software betriebsbereit angeboten. Die Administration liegt vollständig beim Cloud-Anbieter, d.h. dieser kümmert sich um den kontinuierlichen Betrieb und die Wartung.

Cloud-Dienste sind bei privaten Anwendern sehr beliebt, da sie ihnen einen vielfältigen Mehrwert bieten. Beispielweise erlaubt Google Drive[5] das zeitgleiche und gemeinsame Bearbeiten von Dokumenten wie Texten oder Tabellenkalkulationen und eignet sich damit zum gemeinsamen Erstellen von Publikationen (z.B. Konferenzbeiträge), als Materialsammlung oder zum Brainstorming. Dienste wie OwnCloud[6] oder DropBox[7] ermöglichen den Anwendern, Dateien ihres PCs mit der Cloud zu synchronisieren und einfach mit Dritten zu teilen – z.B., wenn der Versand eines großen Dokuments per E-Mail nicht möglich ist.

Bevor sich eine Forschungsinstitution für die Nutzung von Cloud-Computing bzw. einer von einem externen Anbieter betriebenen Cloud-Infrastruktur entscheidet, sollten sorgfältig die folgenden Vor- und Nachteile abgewogen werden:

Vorteile:

- **Kosten:** Die Nutzung der angemieteten Dienste kann erhebliche Einsparungen bewirken. Oft werden nur die Ressourcen abgerechnet, die auch tatsächlich verbraucht werden. Außerdem können angemietete Dienste, Software oder Hardware nur eine begrenzte Zeit angemietet und relativ flexibel gekündigt werden. Darüber hinaus können mit Cloud-Computing die Personalkosten gesenkt werden, da weniger Personal zur Wartung der eigenen IT-Infrastruktur notwendig ist.
- **Aktualität:** Die Ressourcen, insbesondere angebotene Software, sind in der Regel auf dem neusten Stand.
- **Verfügbarkeit:** Die Ressourcen und Dienste sind jederzeit über den Browser und eine Internetverbindung benutzbar – unabhängig davon, wo man diese benutzt. Außerdem können die Daten, Inhalte und Dienste meist auch geräteübergreifend genutzt werden.
- **Flexibilität:** Je nach Bedarf und Nutzungsgrad können Ressourcen angemietet und an die Anforderungen anpasst werden.

Nachteile:

- **Netzzugang:** Der Netzzugang ist bei Cloud-Computing zwingend erforderlich. Die Qualität und die Geschwindigkeit des Internetzugangs beeinflussen die Qualität, mit der Cloud-Dienste genutzt werden können bzw. wie schnell der Datenaustausch zwischen dem Cloud-Betreiber und Nutzer verläuft.
- **Aktualität:** Der vermeintliche Vorteil kann auch als Nachteil gewertet werden, da ständige Aktualisierungen auch Anpassung auf Seiten der Nutzer erfordern können oder die Verwendung weiterer Software unmöglich machen (z.B. aufgrund geänderter Schnittstellen).
- **Datensicherheit:** Bei der Nutzung von Cloud-Computing verliert man die Kontrolle über seine Daten, wenn man diese einem Cloud-Dienstleiter anvertraut. Mangelnder Einfluss seitens des Nutzers auf den Ablageort der Daten in der Cloud tangieren häu-

[5] https://drive.google.com/
[6] http://owncloud.org/
[7] https://www.dropbox.com/

fig Fragen des Datenschutzes. Insbesondere im Falle internationaler Dienstleister, die Daten auf ausländischen Servern vorhalten, unterliegen diese Daten auch anderen rechtlichen Rahmenbedingungen. Weiterhin stellen Clouds ein lohnendes Ziel für Angriffe dar und sind oftmals exponierter als die institutseigene IT-Infrastruktur. Ein gewisses Maß an Vertrauen gegenüber dem Client-Dienstleister und seiner Befähigung zum Schutz der Daten sind daher unumgänglich. Gerade bei Daten mit personenbezogenen Inhalten ist eine Ablage bei Cloud-Anbieter wenig ratsam. Hier ist die Bereitstellung eines eigenen Dokumenten-Servers (s. Abschnitt 5.1) oder der Betrieb einer institutseigenen privaten Cloud-Infrastruktur vorzuziehen, da hier die Regeln selbst festgelegt werden können. Bedingt durch den hohen Aufwand ist dies in der Regel nur für sehr große Organisationen attraktiv (Bräuninger et al. 2012).

- **Anbieterabhängigkeit:** Durch die Verwendung von Cloud-Diensten gerät man zu einem gewissen Grad in die Abhängigkeit vom Cloud-Dienstleister und mithin von seiner Kompetenz, Zuverlässigkeit und Flexibilität. Ein im Bedarfsfall immer möglicher Wechsel des Anbieters sollte jedoch einen entsprechend großen Migrationsaufwand stets mitdenken.

2.5.3 Zugang von außen

Die IT-Infrastruktur kann einen Beitrag dazu leisten, flexible Arbeitszeitmodelle und familienfreundliche Arbeitszeiten zu ermöglichen, beispielsweise für Mitarbeiter mit Kindern. Die Organisation der Heimarbeit sollte so realisiert werden können, dass abwesende Mitarbeiter auf sämtliche für sie relevante Arbeitsressourcen wie Bibliotheksinhalte, Intranet, Forschungsdaten oder fachspezifische Zeitschriften von Zuhause zugreifen können. Um dies zu gewährleisten, sollten entsprechende Schnittstellen zur Integration von privaten und institutseigenen Ressourcen geschaffen werden. Eine schnelle Internetverbindung des Forschungsinstitut ermöglicht bereits prinzipiell den Zugriff auf Anwendungen, Datenbanken und Kommunikationssysteme des Arbeitgebers, eine VPN-Verbindung gewährleistet darüber hinaus eine sichere Verbindung zwischen dem heimischen Rechner und dem Firmennetz.

VPN (Virtual Private Networks) ermöglicht einen verschlüsselten Zugriff auf die lokale IT-Infrastruktur über öffentliche oder private, nicht sichere Verbindungen. Dazu wird ein sogenannter Tunnel zwischen dem Client und dem vertrauenswürdigen internen Netz aufgebaut, der eine verschlüsselte Datenübertragung ermöglicht. Nur entsprechend berechtigte Benutzer können eine VPN-Verbindung aufbauen und nutzen. Vor der Übertragung von Daten werden die beiden Endstellen der aufzubauenden VPN-Verbindung zunächst gegeneinander authentifiziert. Nur im Falle einer erfolgreichen Authentifizierung sind die beiden Endstellen für die eigentliche sichere Datenübermittlung bereit und der Tunnel wird etabliert. Datenpakete werden jeweils vor dem Übertragen vom Sender verschlüsselt und signiert, in ein weiteres Datenpaket gekapselt und dieses dann an den Empfänger gesendet. Dieser wiederum entpackt das Datenpaket, prüft die Signatur, entschlüsselt die Daten und verarbeitet sie dann weiter. Durch Verwendung kryptographischer Verfahren sind Signatur und Verschlüsslung relativ fälschungssicher.[8] Während der Übertragung können Dritte, die die Verbindung

[8] Kryptographische Verschlüsselungs- und Signaturverfahren bieten immer nur eine relative Sicherheit. Ihre Sicherheit hängt zum einen von den verwendeten Schlüssellängen ab, d.h. Verfahren, die

mitlesen, nur die verschlüsselten Datenpakete beobachten, aber nicht deren Inhalt einsehen (Bundesamt für Sicherheit in der Informationstechnik 2014).

Es gibt eine Vielzahl an VPN-Standards und -Protokollen, die verschiedene Übertragungs- und Verschlüsslungstechniken nutzen und von verschiedenen Herstellern implementiert werden (darunter auch Open-Source-Implementierungen wie OpenVPN). Zu den bekanntesten VPN-Protokollen gehören beispielsweise IPSec, SSL-VPN und PPTP – die Auswahl des passenden Protokolls bzw. Produkts ist vom speziellen Anwendungskontext und dem angestrebten Sicherheitsniveau abhängig. Von den verschiedenen Verbindungvarianten ist für Forschungsinfrastrukturen die Client-to-Site die gängigste, um Mitarbeitern auf Dienstreisen oder bei der Heimarbeit den Zugang zu lokalen Ressourcen zu ermöglichen. Die Variante Site-to-Site kann genutzt werden, um kostengünstig Außenstellen über das Internet an die Hauptstelle der Einrichtung anzubinden, da die Notwendigkeit einer Direktleitung zwischen den Standorten entfällt. Zu beachten ist, dass der Datendurchsatz in beiden Fällen von den jeweiligen Anbindungen an das WAN/Internet abhängt.

Abgesehen von VPN-Verbindungen für den Zugriff auf interne Ressourcen steht eine Vielzahl von Videokonferenzsystemen zur Verfügung (z.B. Flashmeeting, Adobe Connect, Skype etc.), die flexible Kommunikation und Kooperation zwischen Mitarbeitern innerhalb und außerhalb der Forschungseinrichtung ermöglichen (Deloitte 2012). Mit Unterstützung von File-Sharing (z.B. Netzlaufwerk-Freigaben über VPN) und webbasierter Projektmanagement-Software können die Mitarbeiter darüber hinaus gemeinsam an Dokumenten und Publikationen arbeiten. Die Realisierung dieser Kommunikations- und Zusammenarbeitsmöglichkeiten erfordert keinen größeren Aufwand und zieht überschaubare Kosten nach sich (Deloitte 2012). Ersparnisse der durch Reduzierung von durch Dienstreisen verursachten Arbeitskosten ergeben sich ggf. auch mit Blick auf eine mögliche räumliche Überbrückung bei dezentraler Lokalisation einzelner Mitarbeiter.

Bei der Nutzung der oben erwähnten Kommunikations- und Kooperationssysteme gilt es jedoch immer zu bedenken, dass die Sicherheit der Kommunikation und Kooperation im virtuellen Raum schwer zu gewährleisten ist. Aus diesem Grund ist eine Verwendung von Tools, die über externe Dienstleister betrieben werden, stets unter diesem Aspekt zu bewerten, sofern im Rahmen einer solchen Kommunikation sensible oder personenbezogene Daten übertragen werden (vgl. Abschnitt 4.2). Die Forschungseinrichtung sollte ihren Mitarbeitern daher klare Richtlinien zur Nutzung solcher Dienste an die Hand geben.

2.6 Datensicherung

Forschungsdaten stellen einen wichtigen Wert einer Forschungsinstitution dar. Daher gehört es zu den wichtigsten Aufgaben einer digitalen Forschungsinfrastruktur, dass diese empiri-

---

vor einigen Jahren noch als sicher galten, sind – bedingt durch die gestiegene Leistungsfähigkeit von Computern – vergleichsweise leicht durch pures „Durchprobieren" („Brute-Force"-Angriff) zu knacken. Zum anderen werden die Verfahren manchmal durch Software-Hersteller fehlerhaft implementiert und diese Fehler können durch Angreifer ausgenutzt werden. Daher ist es ratsam, die Konfiguration des VPN in regelmäßigen Abständen zu überprüfen und ggf. anzupassen. Das regelmäßige Einspielen von Sicherheits-Updates für die VPN-Software sollte als Selbstverständlichkeit angesehen werden. Auch sollten Passwörter und Schlüssel in regelmäßigen Abständen erneuert werden.

sche Basis nicht verloren gehen bzw. verfälscht oder unbrauchbar werden kann (Bundesamt für Sicherheit in der Informationstechnik 2008b).

Hierbei hilft eine durchdachte Strategie zur Datensicherung (engl. *Backup*), die auch die Minimierung möglicher Ausfallzeiten im Falle des Rückspielens gesicherter Daten beinhaltet und insbesondere folgende Probleme adressiert:

- **Datenverlust:** Daten können durch einen Hardware-Defekt, z.B. durch Ausfall einer Festplatte, verloren gehen. Es existieren verschiedene Ansätze, um diesem Problem zu begegnen. So können z.B. RAID-Systeme[9] einen höheren, nicht jedoch einen vollständigen Schutz bieten. Nach dem Austausch einer Festplatte während der Synchronisierung („Rebuild") der Daten kann es beispielsweise zu einem weiteren Ausfall kommen, der einen vollständigen Datenverlust zur Folge hat. Weiterhin schützt ein RAID-System nicht vor Fehlbedienungen. Sowohl die falsche Wartung durch Administratoren als auch das versehentliche Löschen von Daten durch Nutzer kann zu Datenverlust führen. Insofern erhöht ein RAID-System nur die Verfügbarkeit, es ist allerdings kein Ersatz für eine Datensicherung.
- **Datenmanipulation:** Neben der absichtlichen Manipulation durch berechtigte Nutzer (im Falle einer mutwilligen Sabotage), können Daten auch durch Fehlfunktionen in Hardware und Software sowie durch Computer-Viren verändert werden.

Eine Datensicherungsstrategie definiert schriftlich sowohl die Sicherungsmethoden als auch die Sicherungszeiträume und legt diese konzeptuell fest (Doyle 2000). Die Auswahl einer Sicherheitsmethode hängt von vielen Faktoren ab, z.B. von Art und Wert der Daten, aber auch von den Kosten, die Sicherung und Wiederstellung der Daten verursachen. Die Umsetzung einer Datensicherungsstrategie erfordert eine detaillierte Planung, damit sich der Aufwand für den Sicherheitsgewinn lohnt (Eisentraut & Helmle 2013).

Je nach Häufigkeit der Änderung zu sichernder Daten, können verschiedene Sicherungsmethoden eingesetzt werden. Es werden hauptsächlich drei Methoden unterschieden (Bradford & Mauget 2002; ISO/IEC JTC 1/SC 27 2005):

- **Vollständige Sicherung:** Alle Dateien werden komplett kopiert. Diese Sicherung ist die zeit- und übertragungsaufwendigste, da immer alle Daten vollständig und neu übertragen werden müssen. Zwar erfordert sie den größten Speicherbedarf, doch ist sie somit auch die verlässlichste Methode. Ihr Vorzug besteht nicht zuletzt darin, dass die Wiederherstellung der Daten schneller als bei anderen Methoden verläuft, da alle Daten in einem Sicherungssatz vorliegen.
- **Inkrementelle Sicherung:** Bei dieser Sicherung werden nur diejenigen Daten gesichert, die seit der *letzten* Sicherung geändert wurden bzw. hinzugekommen sind. Daher sind Dauer und Speicherplatzverbrauch geringer im Vergleich zur vollständigen Sicherung. Eine vollständige Wiederherstellung auf verschiedene Sicherungssätzen aufgeteilter Daten ist jedoch am zeitaufwendigsten, da die inkrementellen Si-

[9] engl. "Redundant Array of Independent Disks"; manchmal auch "Redundant Array of Inexpensive Disks".

cherungsätze in ihrer jeweils richtigen Reihenfolge zurückgespielt werden müssen. Zusätzlich besteht die Gefahr, dass der Datenbestand eventuell gar nicht oder zumindest nur unvollständig wiederhergestellt werden kann, sofern ein Sicherungssatz unbrauchbar ist.

- **Differenzielle Sicherung:** Bei dieser Methode werden nur diejenigen Daten gesichert, die seit der letzten *vollständigen* Sicherung hinzugekommen sind oder verändert wurden. Die Schnelligkeit dieser Methode ist in der mit der der inkrementellen Sicherung vergleichbar; im Unterschied zu dieser verläuft die vollständige Wiederherstellung schneller, da nur Sicherungssätze der letzten Vollsicherung und der letzten Differenzialsicherung benötigt werden. Auch die Gefahr des Datenverlustes durch einen defekten Sicherungssatz ist geringer. Hingegen ist der Bedarf an Speicherplatz höher als bei der inkrementellen Datensicherung.

In Abhängigkeit von möglicherweise auftretenden Risiken müssen die Backup-Sätze an einem oder mehreren sicheren Speicherorten aufbewahrt werden. Für jeden Speicherort kann dann die Strategie angepasst werden, sodass der Zugriff auf die Sicherungssätze bei Bedarf schnell und einfach verläuft.

Datensicherungen können zu festgelegten Zeitintervallen (täglich, wöchentlich, monatlich) oder auch automatisch nach definierten Ereignissen erfolgen, z.B. nach Übersteigen eines bestimmen Schwellenwerts an Datenänderungen. Oftmals kommen die unterschiedlichen Sicherungsmethoden nach einem festgelegten Plan zur Anwendung. So kann am ersten Wochenende im Monat eine vollständige Sicherung, an den folgenden Wochenenden eine differenzielle Sicherung und an den Werktagen dazwischen eine inkrementelle Sicherung vorgenommen werden.

Ein weiterer Aspekt ist die Aufbewahrungsdauer („retention period“) von Sicherungssätzen. Je länger diese aufbewahrt werden, desto weiter können Daten aus der Vergangenheit rekonstruiert werden, wobei mit steigender Aufbewahrungsdauer der Bedarf an Speicherplatz für die Sicherungssätze steigt. Zum Anlegen von Sicherungssätzen stehen verschiedene Sicherungsmedien zur Verfügung, darunter Bandlaufwerke, Festplatten, Netzwerk- und andere Online-Speicher, die in diesem Kontext allerdings nicht näher erläutert werden sollen.

Eine Datensicherung kann den an sie gestellten Anforderungen nicht gerecht werden, wenn sie nicht regelmäßig überwacht und auf Korrektheit hin überprüft wird. Insbesondere die Wiederherstellung der Daten muss kontinuierlich geprüft werden, um potentielle Datenverluste zeitnah bemerken zu können. Auch kann es vorkommen, dass Dateien nach erfolgreicher Wiederherstellung nicht mehr lesbar sind, wodurch die gesamte Datensicherung wertlos wird. Vorgehensweise und Zuständigkeiten bei der Wiederherstellung müssen genau dokumentiert werden, um Folgefehler zu unterbinden. Die Datensicherungsstrategie erfasst daher in der Regel nicht nur, welche Daten von welchen Systemen mit welchen Verfahren und Zeitabständen gesichert werden und wie lange die Sicherungssätze aufbewahrt werden sollen, sondern auch die Maßnahmen und Regelungen, die im Auftreten eines Notfalls dafür sorgen sollen, dass die gesicherten Daten möglichst schnell und reibungslos wiederhergestellt werden können. Insgesamt ist die Datensicherung ein aufwendiger Prozess, der von der Beschaffenheit der Daten und verfügbaren Ressourcen abhängig ist.

2.7 Nutzung der IT-Infrastruktur für private Zwecke

Generell ist davon auszugehen, dass die Mitarbeiter einer Forschungseinrichtung die Möglichkeiten der dort vorhandenen IT-Infrastruktur nicht ausschließlich für dienstliche Zwecke nutzen werden. Daher sollte es im Interesse der Einrichtung sein, zusammen mit den Mitarbeitern allgemein akzeptierte Verhaltensregeln (z.B. im Rahmen einer verbindlichen Betriebsvereinbarung) für deren Nutzung zu definieren.

Das Ziel einer Betriebsvereinbarung ist es keinesfalls, das Verhalten oder die Leistung der Mitarbeiter zu kontrollieren oder in ihr allgemeines Persönlichkeitsrecht einzugreifen, sondern die Qualität der Arbeit zu fördern und zu schützen und einen Kompromiss zwischen den Interessen der Institution und denen der Mitarbeiter zu schaffen. Die Betriebsvereinbarung zur Nutzung der Forschungsinfrastruktur kann folgende Punkte enthalten:

- **Nutzung von Internet, E-Mail und Telefon für private Zwecke:** Viele Leitlinien und Orientierungshilfen behandeln diesen Aspekt, aber eindeutige Regelungen zur privaten Internetnutzung gibt es nur in wenigen Forschungseinrichtungen, was zu einer unklaren Rechtslage führt (Lepper 2007; Kiesche & Wilke 2011; Die Landesbeauftragte für Datenschutz und Informationsfreiheit 2013).
  Klar definierte Bedingungen verbieten nicht zwangsläufig die private Nutzung dieser Dienste, aber sie können den Rahmen vorgeben, z.B. dass eine private Nutzung nur in geringfügigen Umfang zulässig ist und diese Nutzung den Betriebsablauf nicht stören oder negativ beeinflussen darf. Ferner kann eine Betriebsvereinbarung die Nutzung der Telekommunikationsinfrastruktur für kommerzielle oder gar geschäftliche Zwecke während der Arbeitszeiten untersagen. Ebenso könnten gebührenpflichtige Telefonate ausgeschlossen werden, wobei hier auch technische Maßnahmen der verwendeten Telefonanlage greifen können.
- **Abgrenzung und Kontrolle/Datenschutz und Privatsphäre:** Eine Betriebsvereinbarung sollte deutlich definieren, ob die Nutzung der Infrastruktur für private Zwecke teilweise oder vollständig untersagt ist. Diese Abgrenzung ist jedoch nicht immer einfach. Im Falle von Doktoranden, die zur Abfassung ihrer Dissertation sowohl die IT-Infrastruktur der Institution als auch im Rahmen eines Forschungsprojekts erstellte Forschungsdaten nutzen, ergeben sich Unschärfen zwischen privater und dienstlicher Nutzung. Eine Forschungsinstitution sollte in solchen Fällen ihre Interessen klar definieren. Weiterhin muss geklärt werden, wie die private Nutzung der Infrastruktur unter Berücksichtigung der Belange des Datenschutzes kontrolliert werden kann. Es muss schriftlich fixiert werden, in welchem Umfang private und/oder dienstliche Nutzung von Internet und Telefon kontrolliert bzw. protokolliert werden darf und wie dabei die personenbezogenen Daten geschützt werden können. Dabei ist es wichtig, einen Kompromiss zwischen dem Interesse der Institution nach Kontrolle und der Privatsphäre der Mitarbeiter zu finden. Die frühzeitige Einbeziehung evtl. zuständiger Datenschutzbeauftragter ist hier sinnvoll.

Die Ausarbeitung einer solchen Betriebsvereinbarung zur privaten Nutzung der Infrastruktur beseitigt Unklarheiten sowohl auf Seiten der Beschäftigten als auch der Administration.

2.8 IT-Support, Trouble-Ticket-System

Der IT-Support befasst sich vor allem mit allen Problemen und Fragen, die im Umgang mit den IT-Services und -Systemen auftreten. Der (Teil-)Ausfall der IT-Infrastruktur kann nicht nur hohe Kosten nach sich ziehen, sondern wirkt sich auch auf die Arbeit der ganzen Forschungseinrichtung aus. Neben der Inbetriebnahme und Wartung von Systemen ist eine regelmäßige Wartung der IT-Infrastruktur daher unabdingbar.

Bevor jedoch Probleme behoben werden können, müssen sie entdeckt und eingegrenzt werden. Die steigende Komplexität der Systeme durch Vernetzung, räumliche Verteilung und die zunehmende Heterogenität der Hard- und Softwarekomponenten erlaubt es nicht ohne die notwendigen Spezialfachkenntnisse und eine umfassende Analyse die Ursache für eine Störung einzugrenzen, sie zu identifizieren und die Störung zu beheben. Da die wissenschaftlichen Mitarbeiter einer Forschungseinrichtung im Normalfall für diese Aufgaben weder zuständig noch qualifiziert sind, werden diese Aufgaben daher von einer besonderen Organisationseinheit, dem IT-Service übernommen. Dieser ist auf den Betrieb der IT-Systeme und das Fehlermanagement spezialisiert und garantiert die Verfügbarkeit und Zuverlässigkeit von Systemen und Diensten am Arbeitsplatz (Kruse 2001).

Verschiedene Fehlerdokumentationssysteme oder Störverfolgungssysteme, sogenannte „Trouble-Ticket-Systeme“, erleichtern die zeitnahe Entdeckung und Eingrenzung von Problemen. Darüber hinaus unterstützen sie die Behandlung und Dokumentation von Fehlern in verteilten Netzwerken. Ein Ticket fasst alle Daten zu jeder einzelnen Störung zusammen. Diese können z.B. der Name des Störungsmelders, betroffene Komponente, Störsymptome, Lösung, erfolgte Maßnahmen usw. beinhalten. Die Daten der Tickets werden im System gespeichert, was das Zurückgreifen auf erfolgreich durchgeführte Lösungen bei denselben oder ähnlichen Fehlern erlaubt. Dadurch wird Zeit für eine Lösungssuche eingespart und die Qualität der Störungsbehebung erhöht.

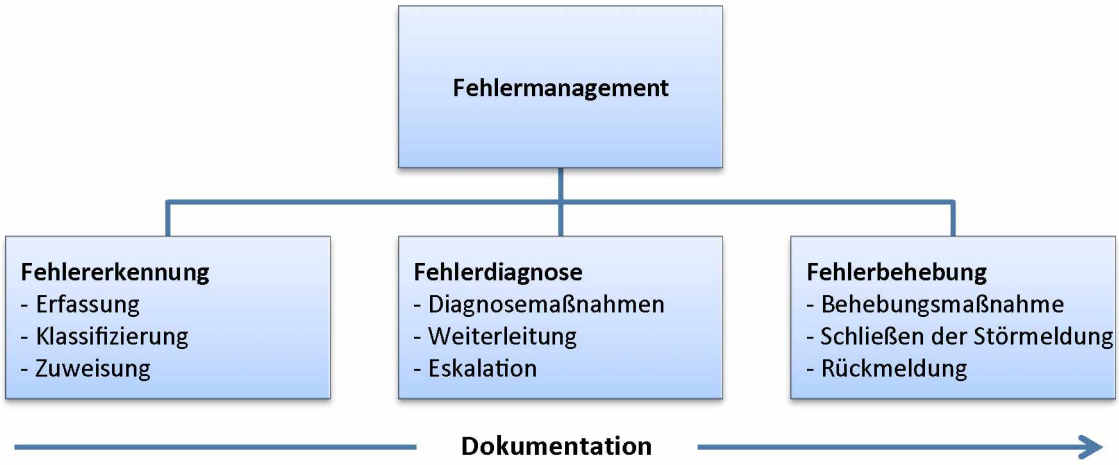


Abbildung 1: Phasen des Fehlermanagements nach (Kruse 2001)

Ein Trouble-Ticket-System kann einerseits Anwendern helfen, Probleme schnell und an eine zentrale Stelle zu melden, bei der sie an die zuständige Person weitergeleitet werden. Anwender bekommen so auch die Möglichkeit, transparente Auskünfte zum Stand der Bearbeitung ihres Problems zu bekommen. Andererseits hilft das Trouble-Ticket-System dem IT-Support zusammen mit der Problemmeldung viele notwendige Informationen wie Bearbeiter, Gruppe, Geräte, Priorität zusammenzufassen, die zur schnelleren Problemlösung beitragen können. Außerdem kann definiert werden, innerhalb welcher Zeit auf das Eintreffen einer Störungsmeldung reagiert werden soll. Im Falle einer Überschreitung wird die Bearbeitungspriorität des Problems automatisch erhöht bzw. die Meldung eskaliert.

Insgesamt führt die Verwendung eines solchen Systems dazu, schneller eine Lösung zu finden und die Service-Leistungen zu verbessern (Nehrenheim 2001; Uhr et al. 2003; Gausemeier et al. 2009), was gerade für Service-orientierte Forschungsinfrastrukturen relevant sein kann.

2.9 Internetauftritt, Content-Management-Systeme (CMS)

Die Präsenz im Internet mittels einer Webseite gehört inzwischen zu den Standardprozeduren eines jeden Forschungsinstituts. Der Verwaltungsaufwand einer solchen Webseite wächst mit dem Bedarf nach der Darstellung aktueller Ereignisse und Forschungsvorhaben oder dem Zugang zu Forschungsdaten und -ressourcen. Zwar wird bei der Gestaltung der Webseite üblicherweise auf ein einheitliches Corporate Design Wert gelegt, doch pflegen einzelne Fachabteilungen oder Projekte ihre Inhalte oftmals selbständig ein, sodass der traditionelle Weg der manuellen Bearbeitung der HTML-Seiten schnell an seine Grenzen stößt.

Abhilfe kann ein Content-Management-System (CMS) schaffen. Besondere fachliche bzw. technische Kenntnisse sind nur bei der ersten Einrichtung des CMS sowie bei der speziellen Anpassung an die Bedürfnisse der Forschungseinrichtung erforderlich. Die Auswahl eines passenden CMS berücksichtigt je nach den individuellen Anforderungen eines Instituts sowohl technischen Aspekte, wie zu unterstützende Betriebssysteme, Wartungsfreundlichkeit und Anpassbarkeit, als auch die redaktionellen Anforderungen, wie Unterstützung von Publikationsprozessen bzw. Gruppen, (Web-)Gestaltungsmöglichkeiten etc. Neben einer Reihe von Open-Source-CMS, wie TYPO3,[10] Joomla[11] oder Drupal,[12] stehen auch kommerzielle Lösungen zur Verfügung. Auch das Mieten einer Hosting-Lösung inkl. CMS-Betrieb durch einen Dienstleister ist möglich, der beispielsweise für alle technischen Aspekte verantwortlich ist (inkl. Aktualisierung von Server und CMS), so dass auf die Forschungsinstitution ausschließlich die redaktionellen Arbeiten entfallen.

2.10 Infrastruktur-Verbünde

Verschiedene Dienste bzw. Komponenten einer Forschungsinfrastruktur, die einen Mehrwert für Mitarbeiter einer Forschungseinrichtung bieten, ergeben sich aus der Teilnahme an Infrastrukturverbünden. In solchen Kooperativen wirken verschiedene nationale und internationale wissenschaftliche Einrichtungen mit, um ihren Nutzern z.B. den einfachen Zugang

[10] http://www.typo3.org
[11] http://www.joomla.org/
[12] https://drupal.org/

zum Internet zu ermöglichen. Diese Verbünde werden meist durch die National Research and Education Networks (NRENs) koordiniert. In Deutschland wird diese Aufgabe durch den DFN-Verein übernommen, sodass es für die Forschungseinrichtungen einen nationalen Ansprechpartner gibt.

Im Folgenden werden die beiden Infrastruktur-Verbünde „eduRoam“ und „DFN-AAI“ beschrieben.

2.10.1 eduRoam

Das Ziel des Verbundprojekts eduRoam[13] („education roaming”) ist die Einrichtung einer globalen Roaming-Infrastruktur für Mitglieder akademischer Einrichtungen. Dazu kooperieren weltweit verschiedene Hochschulen und Forschungseinrichtungen. Der eduRoam-Verbund ermöglicht es den Mitgliedern einer teilnehmen Einrichtung, beim Aufenthalt an einer anderen teilnehmenden Einrichtung schnell, einfach, und vor allem ohne zusätzliche Anmeldung, einen Internetzugang per WLAN unter Verwendung der Zugangsdaten der eigenen Heimateinrichtung zu nutzen. Unabhängig vom weltweiten Standort entstehen für die Nutzer von eduRoam keine zusätzlichen Kosten.

Im eduRoam-Verbund, welcher kontinuierlich ausgebaut wird, sind zurzeit mehrere Tausend Einrichtungen in aktuell über 60 Ländern vertreten. In Deutschland wird eduRoam über das DFNRoaming des DFN koordiniert und in enger Kooperation mit den Rechenzentren der an diesem Wissenschaftsnetz angeschlossenen Einrichtungen aufgebaut. In Deutschland ist DFNRoaming bzw. eduRoam an über 750 Standorten verfügbar.

Der WLAN-Zugang über eduRoam erfolgt grundsätzlich über eine gesicherte Verbindung, d.h. sowohl die Authentifizierung als auch die Datenübertragung sind verschlüsselt und können somit von Dritten nicht mitgelesen werden. Für die Authentifizierung im WLAN wird der Standard IEEE802.1x verwendet, der die Zugangsdaten eines Teilnehmers nur mit dem Authentisierungsdienst seiner Heimateinrichtung austauscht. Der Gastgeber erfährt je nach Konfiguration des Endgeräts nur den Namen der Heimateinrichtung, nicht jedoch das Zugangspasswort des Teilnehmers.

Ungeachtet der Verschlüsselung der Verbindung sollten Nutzer nichtsdestotrotz auf Sicherheitsaspekte achten. An verschiedenen eduRoam-Standorten können verschiedene Verschlüsselungsverfahren für das WLAN zum Einsatz kommen. Daher ist es empfehlenswert, sich gegebenenfalls vor Ort zu informieren. An einigen Standorten wird noch ausschließlich das inzwischen überholte WPA/TKIP Verschlüsselungsverfahren eingesetzt. Im Zweifelsfall kann es also für Nutzer durchaus sinnvoll sein, eine VPN-Verbindung (s. Abschnitt 2.5.3) zur Heimateinrichtung aufzubauen, bevor sensitive Daten übertragen werden.

Der größte Vorteil von eduRoam besteht für Nutzer darin, dass es mit nahezu jedem WLAN-fähigem Endgerät an allen teilnehmenden Einrichtungen und Organisationen ohne größeren Aufwand kompatibel und im Normallfall keine spezielle Software zur Verwendung notwendig ist. Auf dem Endgerät wird lediglich ein Betriebssystem benötigt, das die Unterstützung

[13] https://www.eduRoam.org/

für das Authentifizierungsverfahren 802.1x [14] sowie für die Verschlüsselungsverfahren WPA/AES bzw. WPA/TKIP bietet. Des Weiteren sollte der Benutzer über eine individuelle Benutzerkennung bei der Heimateinrichtung verfügen. Da für akademische Gäste oftmals kein individueller Netzzugang bereitgestellt und verwaltet werden muss, reduziert sich der Arbeitsaufwand für die Administratoren des Gastgebers durch die Bereitstellung von eduRoam spürbar. Weiterhin können Administratoren den Netzzugang über eduRoam leicht derart gestalten, dass Gäste nicht auf institutsinterne Netzwerkbereiche zugreifen können und damit die Sicherheit der eigenen Infrastruktur nicht kompromittiert werden kann. Generell ermöglicht der Netzzugang über eduRoam Nutzern in der Regel nur Zugriffe auf das Internet; für die Verwendung geschützter Ressourcen seiner Heimatseinrichtung ist der Nutzer weiterhin auf den Einsatz von VPN angewiesen.

2.10.2 Authentifizierungs- und Autorisierungs-Infrastruktur (DFN-AAI und eduGAIN)

Die Nutzung einer Vielzahl von externen Diensten bzw. Ressourcen für die Forschung wie Datenbanken, Verzeichnissen, Portalen und insbesondere elektronischen Diensten von Verlagen ist in wissenschaftlichen Institutionen inzwischen gängige Praxis. Oftmals hat der Dienstanbieter ein Interesse daran, dass sein Angebot nur einem begrenzten Kreis von Berechtigten zur Verfügung steht und die Nutzung durch einzelne Benutzer nachvollziehbar ist. Es ist daher oftmals notwendig, dass der Benutzer eine dienstspezifische Zugangskennung beim Dienstanbieter erhält, auf deren Basis dann die Zugangsbeschränkungen realisiert werden, d.h. die Authentisierung und Autorisierung der Benutzer werden durch den Dienstanbieter durchgeführt. Um externe Dienste nutzen zu können, müssen die Benutzer jeweils Zugangsdaten bei den entsprechenden Dienstanbietern beantragen. Nicht nur ist dieses Jonglieren mit vielen verschiedenen dienstspezifischen Zugangskennungen umständlich, auch die Administratoren der Dienstanbieter müssen Zeit für entsprechende Verwaltungsaufgaben aufbringen. Weiterhin möchten viele Anbieter ihre Dienste nur Benutzern aus dem akademischen Umfeld, anbieten. Verlässt ein Nutzer seine Universität bzw. Forschungseinrichtung, sollte der Zugriff auf die Dienste mit seinen Anmeldedaten nicht mehr möglich sein. Mit der Vergabe von dienstspezifischen Zugangskennungen kann diese Anforderung nur schwer oder gar nicht gelöst werden.[15]

Eine Lösungsmöglichkeit für dieses Problem ist eine Authentifizierungs- und Autorisierungs-Infrastruktur (AAI) mit zwei verschiedenen Rollen des Identity Providers (IDP) und des Service Providers (SP). Ein Identity Provider verwaltet Identitäten, d.h. im weitesten Sinne Zugangskennungen, während ein Service Provider Dienste und Ressourcen Dritten bzw. den eigenen Nutzern bereitstellt. Auf diese Weise ermöglicht beispielsweise ein Verlag den Zugang zu Volltexten wissenschaftlicher Publikationen. Normalerweise stellt jede teilnehmen-

---

[14] In den meisten Fällen werden die Extensible Authentication Protocol (EAP) Methoden PEAP oder PAP verwendet, die von den meisten Betriebssystemen bereitgestellt werden.

[15] Es bestünde beispielsweise die Möglichkeit Benutzerkonten mit einem Verfallsdatum zu versehen, so dass die Benutzer sich regelmäßig zurückmelden müssen, wenn sie weiterhin den Dienst nutzen wollen. Im Rahmen dieser Rückmeldung ließe sich auch die akademische Affiliation überprüfen. Allerdings sind solche Verfahren sowohl für Benutzer als auch für die Dienstanbieter mit viel Aufwand verbunden.

de Forschungseinrichtung einen Identity Provider für ihre eigenen Benutzer bereit. Verschiedene Identity und Service Provider können sich zusammenschließen und bilden dann eine sogenannte AAI-Föderation.

Möchte ein Benutzer einen Dienst in Anspruch nehmen, wird die Authentisierung des Benutzers, d.h. das Prüfen seiner Zugangskennung, durch den Identity Provider der entsprechenden Forschungseinrichtung durchgeführt. Als erster Schritt in dem so ablaufenden Workflow wird bei der Aktivierung eines entsprechenden Dienstes des Service Providers die institutionelle Zugehörigkeit eines Nutzers überprüft. Ist diese Authentisierung erfolgreich, wird der Benutzer an den Identity Provider seiner Forschungseinrichtung weitergeleitet und meldet sich dort mit seinen Zugangsdaten an. nach geglückter Anmeldung, wird der Benutzer wieder an den Service Provider verwiesen und ist fortan eingeloggt. Ein Identity Provider kann neben der Information über eine erfolgreiche Authentifikation noch weitere Informationen, sogenannte Attribute, an den Service Provider übermitteln. Diese Attribute können z.B. den Status des Benutzers in der Einrichtung (Student, Mitarbeiter, etc.) enthalten und in die Autorisierungsentscheidung eines Service Providers dann mit einfließen, wenn ein Dienst nur von Mitarbeitern der eigenen Organisation genutzt werden darf.

Der Vorteil einer AAI besteht darin, dass innerhalb der AAI-Föderation Benutzer keine dienstspezifischen Zugangsdaten brauchen. Sie können sich bei allen teilnehmen Service Providern mit ihren Zugangsdaten über den Identity Provider ihrer Forschungseinrichtung anmelden und diese Dienste nutzen. Wenn ein Benutzer seine Forschungseinrichtung verlässt und sein Benutzerkonto gesperrt bzw. gelöscht wurde, verliert er automatisch den Zugang zu allen Diensten innerhalb der Föderation.

Mit der DFN-AAI[16] betreibt der DFN-Verein einen Dienst, dessen Ziel der Aufbau und Betrieb einer nationalen Authentifizierungs- und Autorisierungs-Infrastruktur (AAI) ist. Diese AAI-Föderation ist vorwiegend für die Nutzung im Bereich Forschung und Lehre vorgesehen (Gietz et al. 2006; Borel et al. 2009).Der DFN-Verein tritt beim DFN-AAI als zentraler Vertragspartner auf, der zwischen den verschiedenen Föderationsmitgliedern[17] vermittelt. Dabei werden die organisatorischen und technischen Rahmenbedingungen festgelegt und ein für alle Teilnehmer der DFN-AAI-Föderation rechtlich verbindlicher Vertragsrahmen für den Austausch von Nutzerinformationen vereinbart. Dadurch sind die Rechte und Pflichten der verschiedenen teilnehmenden kommerziellen und nicht-kommerziellen Partner[18] vertraglich festgeschrieben und das für einen solchen Dienst notwendige Vertrauensverhältnis ist gewährleistet. Rein technisch wird die DFN-AAI durch eine auf SAML2 basierte Softwareinfra-

[16] https://www.aai.dfn.de/

[17] Liste der Teilnehmer: https://www.aai.dfn.de/verzeichnis/teilnehmer/

[18] Kommerzielle Partner in diesem Kontext sind meist Verlage, die über eine AAI abgesichert Zugriff auf lizensierte Volltexte aus ihrem Programm bieten. Ein weiter bekannter Partner ist Microsoft mit seinem DreamSpark-Programm (https://www.dreamspark.com), das Teilnehmern von Einrichtungen den Zugang zu verschiedenen Software-Produkten ermöglicht; natürlich davon Abhängig, welche (Lizenz-)Vereinbarungen die Einrichtung mit Microsoft getroffen hat.

struktur realisiert, die bei vielen Teilnehmen durch die Software-Komponenten des Shibboleth-Projekts[19] implementiert wird (Borel et al. 2009; DFN-AAI 2010a; Kähler 2010).

Als Ergänzung zu den nationalen AAI-Föderationen bietet das eduGAIN[20]-Projekt ("GÉANT Authorisation INfrastructure for the research and education community") einen föderationsübergreifenden Dienst, der im Rahmen des pan-europäischen Projekts GÉANT2 entwickelt wurde und sich zum Ziel setzt, eine vertrauenswürdige Infrastruktur für den Zugriff auf Ressourcen und Dienste der GÉANT (GN3plus) Partner zu ermöglichen. Dazu werden in eduGAIN verschiedene AAI-Föderationen über die nationalen Grenzen hinweg zusammengeschlossen und bilden eine AAI-Konföderation (Kersting & Rauschenbach 2008). In eduGAIN wurde eine Architektur entwickelt, die eine Interoperabilität zwischen sowohl vorhandenen als auch zukünftigen AAI-Föderationen ermöglicht. Es wurde bei der Konzeption von eduGAIN besonderen Wert darauf gelegt, dass Eingriffe in bzw. Änderungen an den nationalen AAI-Föderationen nicht notwendig sind.

Bereits heute verbindet eduGAIN weltweit verschiedene AAI-Föderationen. Außerdem laufen Verhandlungen mit verschiedenen nationalen AAI-Föderationen um die Konföderation weiter auszubauen. Der DFN-Verein bzw. die DFN-AAI-Föderation ist für Deutschland Partner in eduGAIN und koordiniert diese Aktivitäten.

3. Bibliothek im Kontext Digitaler Forschungsinfrastrukturen

Viele Medien wie Zeitschriften und Bücher werden zusehends nur noch digital angeboten. Auch die Bandbreite der elektronisch verfügbaren Medien nimmt ständig zu. Damit kommt Bibliotheken als Schnittstelle zwischen Print- und Digitalmedien eine Schlüsselrolle zu, die neben den klassischen Aufgaben wie Erwerb und Katalogisierung auch durch die Aspekte Ressourcenmanagement und Verlag charakterisiert wird. Gerade das Ressourcenmanagement zeichnet sich durch die Notwendigkeit der Datensicherung und der langfristigen Verfügbarkeit digitaler Dokumente (vgl. Abschnitt 5.2.3) aus. An vielen Forschungseinrichtungen übernehmen Bibliotheken die Verwaltung hauseigener Publikationen und die Umsetzung von Open-Access-Angeboten (vgl. Abschnitt 4.1.3), indem sie beispielsweise Dokumentenserver (s. Abschnitt 5.1) aufbauen und betreiben und Publikationen darüber erschließen.

Bibliotheken stehen damit vor der Herausforderung, alle erworbenen digitalen Medien, die in der Forschungseinrichtung entstandenen Publikationen sowie andere wissenschaftliche Daten in ein Repositorium aufzunehmen und zu verwalten. Sie übernimmt neben der Rolle des Empfängers und Verteilers von Wissen und publizierten Produkten damit auch die Rolle eines Verlegers (Klotz-Berendes & Schönfelder 2000). Hinzu kommen die Aufgaben der langfristigen Verfügbarhaltung und Verwaltung des elektronischen Dokumentenbestands. Trotz vieler Normierungen und Standards für die Erstellung und Verwaltung elektronischer Texte sind Bibliotheken auf eine enge Zusammenarbeit mit der vor Ort vorhandenen IT-Infrastruktur angewiesen. Hierbei haben außeruniversitäre Forschungseinrichtungen oftmals aufgrund ihrer geringeren Größe einen Vorteil gegenüber Universitäten, da sich Kooperatio-

[19] http://shibboleth.net/
[20] http://www.edugain.org/

nen zwischen Querschnittsinfrastrukturen wie der Bibliothek und der IT-Abteilung leichter bewerkstelligen lassen.

4. Rechtliche Rahmenbedingungen in digitalem Forschungsraum

Jede Forschungseinrichtung ist sowohl als Rezipient als auch als Produzent von Forschungsdaten. Hierzu zählen nicht zuletzt die von einer Forschungseinrichtung veröffentlichten Publikationen, welche wiederum gewissen rechtlichen Rahmenbedingungen unterliegen. Diese umfassen zunächst Fragen des Urheber- und Verwertungsrechts, auf die im Folgenden eingegangen wird, aber auch rechtliche Aspekte der Archivierung, die gesondert in Abschnitt 5.2.1 behandelt werden.

4.1 Urheber- und verwertungsrechtliche Fragen

Die Erstellung von Forschungsdaten baut häufig auf bereits vorhandenen Ressourcen auf. So wird am Beispiel der germanistischen Sprachwissenschaft bei der Erstellung von (Text-)Korpora häufig auf bereits bestehende Primärdaten zurückgegriffen. In diesem Fall müssen urheber- und verwertungsrechtliche Fragen im Vorfeld geklärt werden – zu beachten sind dabei unterschiedliche Rechtsnormen auf nationaler und internationaler Ebene.

Das deutsche Urheberrecht legt fest, dass Werke der Literatur, Wissenschaft und Kunst (inkl. Software) in ihrer Form (nicht in ihrer Idee) geschützt sind, sofern sie eine individuell erkennbare Schöpfungshöhe aufweisen. Die Rechte an einem Werk hält ausschließlich der Urheber. Darunter fallen das Recht der Verwertung in körperlicher Form, der Vervielfältigung, der Verbreitung, der öffentlichen Wiedergabe und der Erlaubnis der Bearbeitung (z.B. in Form von Annotationen, die einem Primärtext hinzugefügt werden).

Urheberrechte erlöschen erst 70 Jahre nach dem Tod des Urhebers (bei mehreren Miturhebern 70 Jahre nach Tod des längstlebenden Miturhebers) und sind in Deutschland (im Gegensatz zum anglo-amerikanischen Rechtssystem) abgesehen von Vererbung nicht übertragbar. Allerdings kann ein Urheber Dritten (beispielsweise einem Verlag oder einer Forschungsinstitution) einfache oder ausschließliche Nutzungsrechte einräumen.

Während das anglo-amerikanische Rechtssystem den Begriff des „Fair Use“ kennt, worunter bestimmte nichtautorisierte Nutzungsformen zum Zwecke der der öffentlichen Bildung und der Anregung geistiger Produktionen fallen, gibt es in Deutschland sowohl für Privatpersonen als auch für die wissenschaftliche Nutzung sogenannte Urheberrechtsschranken, die Ausnahmen vom Urheberrecht darstellen (§§ 44a-63a UrhG). Sie erlauben und regeln im Rahmen von Forschung und Lehre das Zitieren zur Erläuterung des Inhalts eines eigenen Werkes (unter Angabe der Quelle), sowie die öffentliche Zugänglichmachung kleiner Teile eines Werkes oder einzelner Zeitungen- und Zeitschriftenbeiträge „für einen bestimmt abgegrenzten Kreis von Personen für deren eigene wissenschaftliche Forschung“ – z.B. die Mitarbeiter eines Forschungsinstituts oder -teams (§52a UrhG). Zu beachten ist, dass nach Hoeren (2013, S. 154) Mitarbeiter einer offenen Forschergruppe nicht mehr unter diese Definition fallen dürften und §52a UrhG, auch wenn er bereits mehrfach verlängert wurde (aktuell

bis Ende 2014), zunächst nur befristet gültig ist – eine endgültige Rechtsvorschrift steht noch aus.

Zusammengefasst bedeutet dies, dass eine Forschungseinrichtung, die beispielsweise Annotationen auf Basis von durch Dritte erstellten Daten durchführt, diese zunächst lizensieren muss – auch vor dem Hintergrund einer späteren Veröffentlichung der so modifizierten Daten.[21]

4.1.1 Forschungsdaten

Als Anbieter wissenschaftlicher Daten (z.B. Datenbanken), die frei von Rechten Dritter entwickelt wurden, sind Forschungseinrichtungen in einer ungleich komfortableren Position – können sie doch die Art und Weise der Lizenzierung selbst bestimmen und damit z.B. auch die weitere Bearbeitung (im Sinne einer Korrektur oder Fortführung) ihrer Forschungsergebnisse zulassen oder ausschließen.

Als eine sehr liberale und gleichzeitig gut verständliche Lizenz für Forschungsdaten, die im Rahmen öffentlich geförderter Forschungsprojekte entwickelt werden, hat sich in den letzten Jahren die Creative Commons License (CC) herauskristallisiert. CC gehört zu den freien Lizenzen (ähnlich wie die BSD License[22] oder die GNU General Public License[23] für Software) und wurde speziell für kreative Werke entwickelt. Dabei ist zu beachten, dass freie Lizenzen keine Alternative zum aktuellen Urheberrecht sind, sondern im bestehenden Urheberrechtsrahmen eingesetzt werden können, um einige der dringendsten Fragen in Bezug auf digitale Werke zu lösen.

Während das klassische Urheberrecht nur das traditionelle „Alle Rechte vorbehalten" kennt, ermöglicht CC dem Lizenzgeber, sein Werk unter bestimmten Auflagen zu verbreiten. Vereinfacht ausgedrückt besteht die Lizenz aus einem Baukastensystem, das als Auflagen die Namensnennung (CC BY, Minimalanforderung), den Ausschluss einer Bearbeitung (CC BY-ND) sowie die Weitergabe unter gleichen Bedingungen (CC BY-SA) oder den Ausschluss kommerzieller Nutzung (CC BY-NC) kennt.[24] Die seit November 2013 aktuelle Version 4 hat verständlicher strukturierte Lizenztexte und berücksichtigt Datenbankrechte, für die es in Europa gesonderte Rechtsvorschriften gibt (anders als in den USA).

Eine gute Hilfestellung zur Lizensierung von Forschungsdaten kann auch der modulare Baukasten (im Original als „Waschzettel" bezeichnete) mit verschiedenen Lizenzkategorien des CLARIN-Projektverbunds darstellen (eine ausführliche Beschreibung findet sich in Oksanen *et al.*, eine Erweiterung des Modells in Kupietz & Lüngen, 2014).

Neben diesen Ausführungen sind weiterhin folgende rechtliche Aspekte bei der Aufbewahrung von Forschungsdaten zu beachten:

---

[21] Eine technische Möglichkeit, diese bisweilen unklare rechtliche Situation zumindest zu entschärfen, kann in der Trennung von Primärdaten und Annotation liegen (Standoff-Annotation).
[22] http://opensource.org/licenses/bsd-license.php
[23] http://www.gnu.org/licenses/#GPL
[24] Gerade der Ausschluss kommerzieller Nutzung kann unerwünschte Nebeneffekt mit sich bringen, wie (Klimpel 2012) zeigt.

- Archivkopien und Digitalisate analoger Inhalte sind zu Archivzwecken gestattet (§ 53 Abs. Satz 1 Nr. 2 UrhG), sofern sie der Sicherung des vorhandenen Bestandes dienen, als Vorlage ein eigenes Werkstück, jedoch kein Datenbankwerk zugrunde liegt, keine wirtschaftlichen Zwecke verfolgt und technische Kopierschutzverfahren am Original nicht umgangen werden.
- Ein Harvesting von Internetangeboten, also das automatisierte und möglicherweise regelmäßige Abrufen derselben, ist in Deutschland unabhängig vom angewandten Verfahren ohne Einwilligung des Urhebers unzulässig. Andere Rechtsgebiete haben abweichende Regelungen – so etwa die Vereinigten Staaten von Amerika, wo von einer prinzipiellen Einwilligung ausgegangen wird, sofern diese nicht ausdrücklich ausgeschlossen wurde.
- Bei der Migration und Emulation von Forschungsdaten zur Formataktualisierung kann es zu geringfügigen Eingriffen in die Integrität des Datenbestandes kommen. Dies ist im Rahmen des Urheberrechts nur dann zulässig, wenn Schöpfungshöhe und informationeller Kernbestand im Vergleich zum veralteten Format unangetastet bleiben. Es greifen die Regelungen zur Vervielfältigung und nicht die zur Bearbeitung. Eine Institution, der das Recht zur Anfertigung von Archivkopieren zusteht, darf ungestraft Formatanpassungen durchführen.
- Das Urheberrecht definiert das Archivieren vornehmlich über den Aspekt der Datensammlung und -aufbewahrung (§ 53 Abs. 2 UrhG), während im Selbstverständnis der Archiveinrichtungen die Zugänglichmachung von Inhalten im Vordergrund steht. Archive, deren Zweckpriorität also die Veröffentlichung von Ressourcen ist, haben prinzipiell kein inhärentes Recht zur Anfertigung von Archivkopien. Es empfiehlt sich also stets die Berücksichtigung dieses Punktes bei der Einholung der Zustimmung des Urhebers.
- Wenn auf Daten ausschließlich zu Archivzwecken von institutseigenen Mitarbeitern zugegriffen wird, entstehen kaum rechtliche Schwierigkeiten. Wenn jedoch dieselben Berechtigten Downloads oder Ausdrucke von geschützten Daten zu wissenschaftlichen Zwecken durchführen (Dokumentenserver), handelt es sich um eine zustimmungspflichtige Vervielfältigung im Sinne des Urheberrechts (§ 53 Abs. 2 S. 1 Nr. 1 UrhG). Auch dieser Aspekt ist in Absprachen mit Urhebern zu berücksichtigen.
- Öffentliche Einrichtungen haben seit Einführung des § 52b UrhG die Möglichkeit, Ressourcen über eigens intern eingerichtete Leseplätze Externen zur Verfügung zu stellen, sofern keine kommerziellen Absichten verfolgt oder bestehende Lizenzvereinbarungen verletzt werden. Eine Entscheidung über eine gänzlich offene Erschließung von Archivinhalten, etwa über die Möglichkeiten digitaler Medien und des Internets, obliegt allein dem Urheber (bzw. dem Inhaber der Verwertungsrechte). Ein Haftungsanspruch gegen ein Archiv entsteht bei Verstoß gegen das Urheberrecht oder die entsprechenden Regelungen zur Verbreitung verfassungsfeindlicher oder pornografischer Schriften. Ob aber ein solcher Verstoß vorliegt, ist grundsätzlich eine Einzelfallentscheidung. Dies gilt insbesondere für entsprechend sensible Stellen in gemeinhin als unanstößig bewerteten Werken – potenziell rechtswidrige Aspekte der Werke Goethes führen kaum Regressansprüche nach sich. Die bloße Vermittlung archivexterner Inhalte sollte demnach deutlich als solche gekennzeichnet und ein Haftungsausschluss formuliert werden.

- Personenbezogene Daten unterliegen dem Bundesdatenschutzgesetz und dürfen grundsätzlich nicht ohne ausdrückliche und nachweisbare Zustimmung der betroffenen Personen erhoben, archiviert, vervielfältigt oder veröffentlicht werden (vgl. Abschnitt 4.2).

4.1.2 Software

Für im Rahmen von Forschungsprojekten entwickelte Software ist CC explizit nicht vorgesehen. Hier ist es sinnvoll, eine der freien Software-Lizenzen zu wählen, etwa die beiden oben schon angeführten GNU General Public License (GPL) bzw. die BSD License oder die MIT License.[25] Während erstere eine echte „copyleft"-Lizenz darstellt, d.h. die so lizensierte Software darf nur unter der gleichen Lizenz modifiziert und weitergegeben werden (inkl. Nennung des ursprünglichen Autoren), ist die BSD License eine „non-copyleft"-Lizenz, die es erlaubt, lizenzierte Software auch unter anderen Lizenzen (und in proprietären Produkten) zu vertreiben.

Generell gilt, dass vor der Freigabe einer Software zu prüfen ist, ob sämtliche Bestandteile selbst entwickelt wurden – sofern auf vorhandene Open-Source-Software aufgesetzt wird, kann die Wahl einer Lizenz bereits eingeschränkt sein. Davon abgesehen besteht natürlich weiterhin die Möglichkeit, Software unter einer restriktiveren Lizenz zu veröffentlichen. Dabei ist allerdings zu beachten, dass so oftmals die Möglichkeit einer aktiven Weiterentwicklung nach Projektende effektiv verhindert wird, solange nicht die ursprünglichen Projektmitarbeiter eine solche betreiben.

4.1.3 Publikationen

Über einen Verlag publizierte wissenschaftliche Arbeiten gehen üblicherweise mit der Einräumung eines (oftmals ausschließlichen) Verwertungsrechts einher. Die Verlage übernehmen dabei sowohl die Herstellung als auch die Verbreitung von Monografien und Zeitschriften und verlangen im Gegenzug entsprechende Kosten für den Bezug, was den Zugang zu wissenschaftlichen Arbeiten teilweise erheblich einschränkt. Darüber hinaus hat der Autor nur in seltenen Fällen die Möglichkeit, das Werk auf anderem Wege (z.B. auf einem Publikations- oder Dokumentenserver seiner Forschungseinrichtung, vgl. Abschnitt 5.1) zugänglich zu machen.

Das Open-Access-Modell[26] kann hier einen Gegenpol bilden, indem Wissenschaftler ihre Publikationen selbstständig über das Internet verbreiten, ohne auf einen Verlag angewiesen zu sein (Degkwitz 2007). Sie sind hierdurch nicht nur in der Lage, ihre Verbreitungs- und Verwertungsrechte zu wahren, sondern auch die Informationsversorgung von der Kostenfrage abzutrennen und der Produktion neuen Wissens damit Vorschub zu leisten „als das traditionelle auf Ausschließlichkeitsrechte aufbauende System." (Spindler 2006, S. I).

Das Konzept hat sich inzwischen im Forschungsbereich etabliert und fördert den für den Rezipienten kostenfreien Zugang zu wissenschaftlichen Informationen und Literatur zum

---

[25] Daneben existiert noch eine nahezu unüberschaubare Anzahl an weiteren Lizenzen. Einen ersten Überblick verschafft die Webseite der Open Source Initiative unter http://opensource.org/.
[26] http://open-access.net/de/startseite/

Zwecke der Bildung und Wissenschaft.[27] Die Open-Access-Bewegung ist in vielen Ländern wie in den USA, Dänemark und Japan stark vertreten, insbesondere die naturwissenschaftlichen Disziplinen profitieren davon. In Deutschland wird dagegen die Diskussion um Open-Access-Publikationen vor allem in den Geisteswissenschaften noch sehr kontrovers geführt (Görl et al. 2011). Die großen Forschungsförderinstitutionen wie die DFG (Fournier 2005), aber auch viele universitäre und außeruniversitäre Institutionen unterstützen die Forderung nach einer Möglichkeit der elektronischen Publikation nach dem Open-Access-Prinzip und unterstützen die „Berliner Erklärung über den offenen Zugang zu wissenschaftlichem Wissen".[28]

Ein Nachteil von Open-Access-Publikationen ist die aktuell oftmals noch geringere Reputation gegenüber Toll-Access-Zeitschriften, wobei üblicherweise beide Publikationsorgane auf das in der Wissenschaft gängige Peer-Review-Verfahren setzen. Zu den Vorteilen gehört dagegen die bereits nachgewiesene höhere Zitationszahl von auch offen zugänglichen Publikationen. Manche Open-Access-Zeitschriften erreichen jedoch sehr hohe Zitierraten und haben sich Spitzenplätze unter den wissenschaftlichen Zeitschriften erobert (Bargheer et al. 2006; Stempfhuber 2009).

In der Praxis haben sich unterschiedliche Open-Access-Strategien etabliert. Zu nennen sind der goldene und der grüne Weg als die beiden Hauptstränge sowie der graue Weg als Mischform.

Auf dem goldenen Weg erfolgt die Erstveröffentlichung einer wissenschaftlichen Publikation und Monografien über einen Open-Access-Verlag, wobei die Finanzierung oftmals unter Zahlung sogenannter Publikationsgebühren durch die Autoren oder deren Forschungsförderern erfolgt (Schmidt 2007). Die Autoren behalten in der Regel die Rechte an ihren Texten. Dagegen zielt der grüne Weg vorrangig auf die Selbstarchivierung von Postprints bereits publizierter wissenschaftlicher Arbeiten in digitalen Repositorien ab (z.B. in frei zugänglichen Online-Archiven oder auf einem institutionellen oder fachspezifischen Dokumentenserver, vgl. Müller & Schirmbacher 2007). Dabei wird die nach siebenjähriger Diskussion im Jahr 2013 beschlossene Einführung eines Zweitverwertungsrechts (§38, 4), das den Autoren nach Ablauf einer Jahresfrist (nach Erstveröffentlichung) die Möglichkeit zur freien Veröffentlichung der Manuskriptfassung einräumt, aufgrund der zu starken Einschränkungen weiterhin als unzureichend angesehen.[29]

Der graue Weg betrifft nur Open-Access-Publikationen, die nicht über Verlage, Zeitschriften und sonstige herkömmliche Vertriebswege veröffentlicht wurden. Hierunter fallen bei-

---

[27] Betont sei hier die Rolle des Rezipienten: einige Open-Source-Publikationen verlangen vom Autoren eine finanzielle Beteiligung, wobei eine durchaus gerechtfertigte Kostenübernahme leider nicht immer mit der Qualität eines Open Access-Journals korreliert.

[28] http://openaccess.mpg.de/3515/Berliner_Erklaerung

[29] Zu den genannten Einschränkungen gehört u. a., dass der Beitrag mindestens zur Hälfte durch öffentliche Mittel gefördert wurde und die Erstveröffentlichung in einem mindestens halbjährlich erscheinenden Periodikum erfolgte. Erfolgen Buchpublikationen ohne Vergütung können sie nach §38, 2 ebenfalls nach Jahresfrist öffentlich zugänglich gemacht werden.

spielsweise Tagungsberichte, Abstract-Sammlungen, Dissertationen und ähnliche Dokumente.

Insgesamt betrachtet bietet Open Access publizierenden Forschungseinrichtungen und Autoren in mehrfacher Weise Vorteile: Auf Rezipientenseite erleichtert das Open-Access-Modell die rechtssichere Nutzung aktueller wissenschaftlicher Forschung, auf Produzentenseite ermöglicht es die eigenen Erkenntnisse einem möglichst breiten Publikum zur Verfügung zu stellen. Insofern können gerade kleinere außeruniversitäre Forschungseinrichtungen hiervon profitieren.

4.2 Persönlichkeitsrechte und Datenschutzrecht

Die Erstellung von Forschungsdaten berührt oftmals Persönlichkeitsrechte, z.B. bei der Aufnahme von Gesprächen. Hier lassen sich teilweise schon anhand der Stimmen einzelne Personen identifizieren, sodass die Einräumung der entsprechenden Rechte zur Veröffentlichung im Rahmen der Forschung durch den Interviewten zwingend notwendig ist. Diese hat schriftlich zu erfolgen und muss entsprechend dokumentiert werden.

Generell gilt, dass für die Sammlung aller personenbezogenen Daten ein Verfahrensverzeichnis nach § 4g i. V. m. §§ 18 und 4e BDSG anzulegen ist, das darüber Auskunft gibt, welche personenbezogenen Daten unter Verwendung welcher automatisierten Verfahren auf welche Weise verarbeitet oder genutzt werden und welche Datenschutzmaßnahmen durchgeführt werden. Üblicherweise werden dazu folgende Punkte dokumentiert:

- Name des Verfahrens (z.B. des Forschungsprojekts),
- Name und Anschrift der verantwortlichen Stelle (inkl. Nennung eines Vertreters),
- Datenschutzbeauftragter,
- Zweckbestimmung und Rechtsgrundlage (hier kann auch ein Verweis auf eine entsprechende vorliegende Einwilligung der Personen, deren Rechte berührt werden, erfolgen),
- betroffene Personen und Daten(-kategorien),
- Empfänger, denen die Daten mitgeteilt werden (hier ist zu unterscheiden zwischen nationalen Empfängern, EU-Staaten und weiteren Drittstaaten),
- Regelfristen für die Löschung der Daten
- Angaben zu den getroffenen Sicherheitsmaßnahmen und deren Angemessenheit.

Generell gilt, dass auch hier der Datenschutzbeauftragte der Forschungseinrichtung frühzeitig zu Rate gezogen werden sollte, sofern Forschungsdaten erhoben werden sollen.

5. Wissensressourcen der wissenschaftlichen Forschung

Neben Kommunikation und Interaktion einzelner Wissenschaftler untereinander dienen Forschungsinfrastrukturen vornehmlich der Erzeugung, Erfassung, Verwaltung, Erschließung und Speicherung unterschiedlichster Ressourcen als Grundlage der wissenschaftlichen Arbeit. Diese Ressourcen reichen von Rohdaten bis hin zu wissenschaftlichen Publikationen. Sicherheit, schneller Zugriff und langfristige Verfügbarkeit dieser Ressourcen bilden die prioritären Handlungsfelder von Forschungsinfrastrukturen. Das aktuelle Kapitel umreißt die wesentlichen Ressourcenarten und zeigt Wege zu einem adäquaten Umgang auf.

5.1 Dokumentenserver

Wissenschaftliche Forschung produziert eine Vielfalt elektronischer Publikationen, deren Vertrieb über kommerzielle Verlage nicht zwangsläufig im Sinne des akademischen Autoren ist (vgl. Abschnitt 4.1.3). Alternativ dienen daher sogenannte Dokumentenserver (Repositorien) dazu, diese Publikationen in der Verantwortung der Forschungsinstitutionen zu sammeln, zu archivieren und zur Verfügung zu stellen (Foster & Gibbons 2005; Bertelmann 2006; Jones 2006). Heute betreibt nahezu jede größere Institution oder wissenschaftliche Einrichtung Dokumentenserver, die sich in institutionelle und disziplinäre Dokumentenservern unterschieden lassen. Institutionelle Dokumentenserver werden folgendermaßen definiert:

> "[...] an institutional repository is a digital archive of the intellectual product created by the faculty, research staff, and students of an institution and accessible to end users both within and outside of the institution, with few if any barriers to access." (Crow 2002, S. 16)

Zu den wesentlichen Ressourcen und wissenschaftlichen Inhalten zählen insbesondere Publikationen, E-prints, Diplomarbeiten, Berichte, Konferenzbeiträge oder Arbeitspapiere. Institutionelle Dokumentenserver werden in der Regel von institutsinternen Bibliotheken (oft in Kooperation mit der ZDV) betrieben. Sie bieten den Mitarbeitern die Möglichkeit, ihre wissenschaftlichen Arbeiten kostenfrei als elektronische Publikationen zu veröffentlichen, die sich im Gegensatz zu den auf disziplinären Dokumentenservern gespeicherten wissenschaftlichen Inhalten nicht auf eine Fachrichtung begrenzen.

Disziplinäre Dokumentenserver sind institutionsübergreifend aufgestellt und enthalten Dokumente eines bestimmten Wissenschaftsgebietes. Sie stehen für die Publikation und Archivierung von fachspezifischen Arbeiten zur Verfügung (z.B. PsyDok[30] für das Fach Psychologie, GiNDok[31] für das Fach Germanistik). Für die Gespräche, die im Zuge der in den Kapiteln 6 und 7 dargelegten Studie mit Vertretern von Forschungsinstituten durchgeführt wurden, standen jedoch institutionelle Dokumentenserver im Vordergrund.

Die Veröffentlichung von wissenschaftlichen Texten auf einem Dokumentenserver bietet eine Alternative zur Verlagspublikation. Durch die Aufnahme in eine solche Speicherumgebung wird den vorgehaltenen Publikationen eine dauerhafte Adresse zugewiesen, sodass sie über nationale und internationale Kataloge, Suchmaschinen sowie andere Nachweisinstrumente erschlossen und aufgefunden werden können; der Server kann so eingerichtet werden, dass die gehosteten Dokumente entweder frei im Internet verfügbar oder aber kostenpflichtig abrufbar sind.

Der Betrieb eines Dokumentenservers ist kostspielig und erfordert sowohl technisches und fachliches Wissen als auch den dauerhaften Einsatz von Institutsressourcen. Für eine Bibliothek selbst ist dieser Aufwand häufig zu hoch. In der Regel arbeitet sie daher beim Betrieb des Servers mit der IT- bzw. EDV-Abteilung eines Instituts zusammen, die üblicherweise die technische Seite der Realisierung eines Repositoriums wie z.B. die Sicherung des Servers und

---

[30] http://psydok.sulb.uni-saarland.de/
[31] http://www.germanistik-im-netz.de/gindok/

der Daten abdeckt, während sich die Bibliothek der Sammlung, der Indizierung und Klassifizierung der abzulegenden Inhalte widmet.

Da für Aufbau und Betrieb eines institutionellen Dokumentenservers verschiedenste Softwaresysteme zur Anwendung kommen, verfügen gerade kleine Forschungseinrichtungen nicht über ausreichende Kapazitäten. Oft fällt die produzierte Menge an zu erwartenden potenziellen Inhalten kleinerer Häuser zu gering aus, um einen eigenen Server ökonomisch zu betreiben, sodass es nützlich sein kann, die Angebote von großen Bibliotheken in Anspruch zu nehmen und wissenschaftliche Volltexte (Konferenzpapiere, Reports, Artikel etc.) in einem vordefinierten Format verfügbar zu machen.

5.1.1 DSpace

DSpace[32] wurde vom *Massachusetts Institute of Technology* (MIT) in Kooperation mit einer kommerziellen IT-Firma 2002 entwickelt. Es handelt sich um eine Software zum Aufbau und Betrieb eines Dokumentenservers, die sowohl für akademische, gemeinnützige als auch kommerzielle Zwecke unter der BSD-Lizenz genutzt werden kann. Für die Weiterentwicklung von DSpace wurde eine gesonderte Föderation gegründet, in der mehrere große Forschungseinrichtungen zusammengeschlossen sind, um die technologische Anpassung des Systems voranzutreiben. Eine ständig wachsende Gemeinschaft von Entwicklern erweitert und verbessert die Software, von der in der Folge auch Fachwissenschaftler profitieren können. Die im Rahmen jährlicher Versionsupgrades erscheinenden Erweiterungen umfassen regelmäßig Anpassungen an geänderte rechtliche Rahmenbedingungen im Publikationswesen, beispielsweise neue Versionen von CC-Lizenzen oder die Möglichkeit, im Rahmen des Zweitverwertungsrechts (siehe Abschnitt 4.1.3) gewisse Fristen für den öffentlichen Zugang zu bestimmten Publikationen zu definieren.

Die Software erlaubt ein hohes Maß an Konfigurierbarkeit, sodass ein Dokumentenserver ohne besonders komplizierte Eingriffe in den Quellcode an die Anforderungen einer Institution angepasst werden kann. Sollten hingegen tiefergreifende Modifikationen an der Software vonnöten sein, erleichtern deren starke Modularisierung und detaillierte Dokumentation sowie Hilfe aus der Entwicklergemeinschaft die Problemlösung.

Neben seiner Eigenschaft, einen einfachen Zugang zu verschiedenen Arten von digitalen Inhalten, z.B. Texten, Bildern, Datensätzen, etc. zu ermöglichen, zeichnet sich DSpace durch ein flexibles und kaskadierendes Rechtemanagement aus. So ist es möglich, Benutzer des Systems verschiedenen Gruppen zuzuordnen, die wiederum Teil übergeordneter Gruppen sein können, wodurch Lese- und Schreibrechte für Dokumente in spezifischen Bereichen des Dokumentenservers detailliert an die Organisationsstruktur einer Institution angepasst werden können.

[32] http://www.dspace.org/

5.1.2 ePrints

ePrints[33] wurde von der Universität Southampton im Jahr 2000 entwickelt. Es bietet einer Einrichtung ein kostenloses Dokumentenarchiv, das unter GPL genutzt werden kann. Die Plattform wurde in der Programmiersprache Perl implementiert und ist derzeit in der Version 3.3.12[34] verfügbar. Die Web-Plattform zur Veröffentlichung von Publikationen bietet Wissenschaftlern die Möglichkeit, auf abgelegte Inhalte zuzugreifen und den Gesamtbestand zu durchsuchen. Das System verfügt über ein Rechtemanagement und garantiert eine serverübergreifende Abfrage von Metadaten.

ePrints kann für verschiedene Arten von Repositorien eingerichtet werden, darunter Forschungsrepositorien, Repositorien für wissenschaftliche Abschlussarbeiten, Datenrepositorien oder Projektrepositorien. Obgleich das System mit dem sogenannten „Self Archiving" so konzipiert ist, dass es Wissenschaftler auch selbst für Eigenpublikationen nutzen können, wird es häufig zum Betrieb von institutionellen Repositorien eingesetzt, wo es den Publikationsprozess für Autoren, Herausgeber und Institutionen gleichermaßen unterstützen kann.

5.1.3 eSciDoc

eSciDoc[35] ist eine eResearch-Umgebung, die ebenfalls als Plattform zur Speicherung und Verwaltung von Forschungsprimärdaten, Analysedaten und Publikationen genutzt werden kann. Sie wurde als gemeinsames Projekt zwischen dem FIZ Karlsruhe und der Max Planck Digital Library (MPDL) entwickelt und im Rahmen der e-Science-Initiative des Bundesministeriums für Bildung und Forschung (BMBF) bis 2009 gefördert. Das System ist vor allem für den Einsatz im wissenschaftlichen Umfeld vorgesehen, um einen uneingeschränkten Zugriff auf Daten, Tools und andere Ressourcen zu ermöglichen.

Ziel des eSciDoc-Projekts war es, einen offenen, sicheren sowie dauerhaften Zugang zu Forschungsdaten und -materialien wissenschaftlicher Institutionen und Forschungsorganisationen zu gewährleisten, um die wissenschaftliche Interaktion und Forschung zu optimieren. Die eSciDoc-Software steht als Open-Source-Angebot unter der CCDL-Lizenz[36] frei zur Verfügung.

5.2 Forschungsdaten und ihre Archivierung

Folgt man der von der DFG vorgegebenen Definition, sind „Forschungsprimärdaten [...] Daten, die im Verlauf von Quellenforschungen, Experimenten, Messungen, Erhebungen oder Umfragen entstanden sind." (DFG 2009, S. 2)

Die zunehmende Digitalisierung der Wissenschaft birgt nicht minder denn analoge Archivierung das Risiko, dass unter hohem finanziellem Aufwand erhobene Daten, welche die Grundlage sowohl für aktuelle als auch künftige Forschung bilden, nach einer relativ kurzen Phase der Auswertung lediglich einzelnen oder kleinen Gruppen zur Verfügung stehen. Sie können möglicherweise schon nach kurzer Zeit nicht mehr gepflegt oder gelesen werden. Im schlimmsten Fall unterliegen die Daten durch technisches bzw. menschliches Versagen oder

---

[33] http://www.eprints.org/
[34] Stand: 31.03.2014
[35] https://www.escidoc.org/
[36] Common Development and Distribution License (http://opensource.org/licenses/CDDL-1.0)

durch die Verwendung proprietärer Formate einem Totalverlust und können für eine wissenschaftliche Nachnutzung nicht mehr wiederhergestellt oder reproduziert werden (Oßwald et al. 2012). Überdies kommt es vor, dass die Daten nach dem Auslaufen finanzieller Förderung von Datenerhebungsprojekten oder aufgrund mangelnder technischer oder fachwissenschaftlicher Nachhaltigkeitslösungen (proprietäre Formate, keine aktive Nutzer- und Entwicklercommunity) nicht vollständig ausgewertet wurden und erst zu einem späteren Zeitpunkt für anschließende Studien wieder relevant und neu analysiert werden müssen. Nicht selten kann zudem nur durch eine kontinuierliche oder nachträgliche Analyse von Forschungsprimärdaten nachvollzogen werden, wie Forschungsergebnisse in der Primärstudie entstanden und begründet worden sind. Die Replizierbarkeit von Forschungsergebnissen ist somit eine zentrale Aufgabe, um im Zweifelsfall wieder auf den Originaldatenbestand zurückzugreifen und die Untersuchung zu reproduzieren.

In diesem Punkt sehen viele Initiativen und Organisationen (z.B. DFG, nestor[37]) in allen Wissenschaftseinrichtungen einen dringenden Handlungsbedarf hinsichtlich der systematischen Sicherung, Archivierung und Bereitstellung von Forschungsprimärdaten, um sie durch Dritte langfristig nutzbar zu halten. Unter den von der DFG definierten „Empfehlungen für eine gute wissenschaftliche Praxis" findet sich unter anderem auch die Anforderung, Forschungsdaten für mindestens zehn Jahre auf geeigneten Datenträgern sicher aufzubewahren (DFG 2009).

Obwohl die Bedeutung der Archivierung und Verfügbarkeit von Forschungsprimärdaten für das wissenschaftliche Arbeiten allgemein bekannt ist, wird dies häufig vernachlässigt. Dies erklärt sich zumeist dadurch, dass Zuständigkeiten und Finanzierung nicht geklärt sind – so verfügen meist nur größere Forschungsverbünde, die eine enge Zusammenarbeit von mehreren Einrichtungen anstreben, über eine Strategie für ein ganzheitliches Datenmanagement (Klump 2010). Zweifellos ist die öffentliche und frei verfügbare Zugänglichmachung erhobener Daten nach Abschluss eines Forschungsprojekts mit erheblichem personellem und finanziellem Aufwand verbunden. Dennoch ist eine nachhaltige Archivierung und barrierefreie Veröffentlichung von Forschungsdaten unabdingbar, um anderen Wissenschaftlern die Möglichkeit zu geben, die Ergebnisse nicht nur nachzuvollziehen, sondern auch in eigenen Studien aufzugreifen, zu ergänzen oder in anderen Kontexten zu nutzen. Der erste Schritt zu einem überzeugenden Datenmanagement muss daher eine Gewichtung des Aufwandes der Archivierung von Daten gegenüber ihrer Sicherungswürdigkeit sein. Entscheidend ist hier, inwiefern die Daten über die konkreten Forschungsfragen im Kontext ihrer Erhebung hinaus von Bedeutung sind. Da sich zukünftige Trends und Methoden der Wissenschaft jedoch nur sehr schwer voraussagen lassen, lässt sich ein „Wert" von Forschungsdaten nur sehr schwer abschätzen. Dieses Kriterium für die Archivierung sollte auf reine Forschungsdaten deshalb nur ausgesprochen behutsam angewandt werden. Hilfreich ist es, den Kontext zu berücksichtigen, in dessen Rahmen die Daten entstanden sind. Sollten die Daten unter nicht replizierbaren einmaligen Rahmenbedingungen gesammelt worden sind, ist eine Archivierung geboten. Ein weiteres Kriterium ist die Abdeckung des Themengebiets, dem die Daten zugeführt werden sollen. Wenn betroffene Gebiete Forschungslücken aufweisen oder die fragli-

[37] http://www.langzeitarchivierung.de/

chen Forschungsdaten Ausnahmecharakter haben, hat eine Archivierung besondere Priorität. Bei jeder Datenerhebung entstehen zudem neben den Primärdaten auch zahlreiche Kontextdaten wie Postskripta, Fallbeschreibungen und -auswertungen (Kretzer 2013). Auch hier ist eine Priorisierung und Hierarchisierung vorzunehmen.

Zweitens ist zu entscheiden, in welchem Format die Daten gespeichert und aufbewahrt werden sollen, sodass sie nicht nur für eigene Zwecke nachhaltig genutzt, sondern auch in anderen Anwendungskontexten möglichst reibungslos eingesetzt werden können.

Drittens müssen Daten mit Metadaten versehen werden, die später als Grundlage der Suche und Verwaltung dienen. Metadaten können Information über die die Datenerhebung motivierende Studie selbst, über Zeitpunkt und Ort der Datenerzeugung oder Inhalt der Daten umfassen. Die Erstellung und Pflege der Metadaten sollte unter besonderer Sorgfalt aufgebracht werden, ermöglichen sie doch nachträglich und kontinuierlich die Daten transparent, interpretierbar und auffindbar zu halten. Beispielsweise kann durch die Nutzung administrativer Metadaten jede Bearbeitung oder Modifikation am Datensatz dokumentiert und nachverfolgt werden (Kretzer 2013). Nachträgliche Kosten und Arbeitsinvestitionen im Versäumnisfalle erübrigen sich bei konsequenter und systematischer Anreicherung als Teil der Langzeitarchivierungsstrategie eines Hauses.

geben sich eine Reihe von Anforderungen hinsichtlich des Datenformats, bevor sie ins Datenarchiv aufgenommen werden. Die Dokumentation von Forschungsdaten spielt hierbei eine zentrale Rolle und gibt für spätere Interpretationsversuche entscheidende Hilfestellungen. Die Erhebungsbedingungen und -kontexte sowie die angewandten Forschungsinstrumente, welche die Erzeugung der Daten charakterisierten, zu dokumentieren und festzuhalten, ist der entscheidende Schritt für eine langfristige Nutzbarkeit. In einer solchen Dokumentation werden inhaltliche, methodische und technische Charakteristika der Studie beschrieben. Zusätzlich zur Dokumentation werden dann Metadaten erstellt (National Information Standards Organization 2004; Jensen et al. 2011). Für eine umfassendere Betrachtung technischer, rechtlicher und organisatorischer Fragen sei auf (Klimpel & Keiper 2013) verwiesen.

5.2.1 Rechtliche Fragen der Archivierung

Die DFG empfiehlt Daten in einer Form bereitzustellen, die es Forschern bei Bedarf erlaubt, Ergebnisse zu replizieren oder auch unter neuen Fragestellungen zu bearbeiten (DFG 2009). In diesem Sinne sollten die Forschungsprimärdaten nicht nur für möglichst unbegrenzte Zeit aufbewahrt, sondern auch öffentlich gemacht werden. Murray-Rust führt in diesem Zusammenhang den Begriff „Open Data“ ein, der einen offenen Zugang zu Forschungsdaten beschreibt, und an die Terminologie von Open Access (vgl. Abschnitt 4.1.3) und Open Source (vgl. Abschnitt 4.1.2) angelehnt ist:

„Open Data (OD) is an emerging term in the process of defining how scientific data may be published and re-used without price or permission barriers.“ (Murray-Rust 2008)

Frei zugängliche Forschungsdaten und andere Quellenmaterialien können den Forschungsprozess beschleunigen und qualitativ zu verbessern, indem die Nachprüfbarkeit und Nach-

nutzbarkeit von Daten gewährleistet wird (Pfeiffenberger & Klump 2006; Pampel et al. 2010).

Der Ansatz von offen zugänglichen Daten bringt allerdings neue Herausforderungen rechtlicher Art mit sich, die von Forschungseinrichtungen stets miteinzubeziehen und oft nicht eigenständig zu lösen sind (Neuroth et al. 2010). Auch bei der digitalen Langzeitarchivierung greifen Aspekte des Urheberrechts, das aufgrund seiner hohen Komplexität auf breite Kritik stößt. Weiterhin gelten insbesondere die in Abschnitt 4.1.1 aufgeführten rechtlichen Aspekte zur Archivierung von Forschungsdaten.

Sofern eine Forschungseinrichtung noch keine Forschungsprimärdatenstrategie erarbeitet hat, empfiehlt es sich dringend, dies nachzuholen und die Infrastruktur für Forschungsprimärdaten anzupassen. Die Größe des Aufwands hängt von der Art der Daten ab, die aus den Forschungsprojekten zur Verfügung gestellt werden, und welche ergänzenden Arbeiten wie Erstellung von Metadaten und Prüfung der Authentizität Daten zu ihrer Anreicherung benötigt werden.

5.2.2 Eine Frage des Formats

Die digitale Verarbeitung von Forschungsdaten impliziert deren Speicherung in einem Datenformat. Dieses kann durch die zur Erstellung oder Verarbeitung herangezogene Software beeinflusst werden, vorzugsweise allerdings sollten Formate verwendet werden, die von mehr als von mehr als einem einzigen Anwendungsprogramm geöffnet werden können. Dies ist vorrangig dann möglich, wenn die Spezifikation des Datenformats offen vorliegt, sodass es unterschiedlichen Anbietern möglich ist, Import- und Exportfilter zu erstellen und es der Forschungseinrichtung ermöglicht, ein Vendor-Lock-In (also die Abhängigkeit von einem Hersteller) zu vermeiden. Ebenso vorteilhaft ist es, wenn das Datenformat nicht in einem Binärformat vorliegt, sondern in einer textuellen Repräsentation, sodass sie sich mit beliebigen Texteditoren bearbeiten lassen. Gerade in der Forschungslandschaft haben auf XML basierende Formate einen breiten Zuspruch erfahren. So gibt es standardisierte Auszeichnungssprachen für Sprachkorpora (z.B. XCES, Ide et al. 2000) oder das Linguistic Annotation Framework (ISO/TC 37/SC 4 2012) bzw. für literarische Texte (z.B. Text Encoding Initiative, Burnard & Bauman 2014).

Sofern die bereits vorhandenen Annotationsformate nicht den eigenen Ansprüchen genügen, sollten vollständige Eigenentwicklungen vermieden werden, sondern Adaptionen der bestehenden Spezifikationen angestrebt werden, z.B. durch Mitarbeit in den entsprechenden Gremien oder durch oftmals zur Verfügung stehende Schnittstellen zur Modularisierung (so hat das IDS beispielsweise eine modifizierte Fassung des XML-basierten Corpus Encoding Standards auf Basis der TEI P5 entwickelt, vgl. Lüngen & Sperberg-McQueen 2012).

So annotierte Daten lassen sich deutlich einfacher archivieren, da sie nicht von spezifischer Software (oder gar Softwareversionen) abhängig ist, die in einem solchen Fall mitarchiviert werden müsste. Generell ist darauf zu achten, dass die zugehörige Dokumentgrammatik für XML-basierte Formate offen vorliegt (und auf dem aktuellen Stand ist) und verwendete Versionen ebenfalls archiviert werden. Eine solche Dokumentgrammatik kann mit Hilfe unterschiedlicher Grammatikformalismen erstellt werden (die geläufigsten sind DTD, XSD und RELAX NG), beschreibt formal eindeutig die entsprechende Auszeichnungssprache und er-

möglicht so die maschinelle Überprüfung auf Validität der so annotierten Daten – ein gerade bei großen Datensammlungen nicht zu unterschätzender Vorteil.[38] Auch selbst entwickelte bzw. adaptierte Stylesheets zur Verarbeitung sollten mit archiviert werden, ebenso wie eine zusätzliche menschenlesbare Dokumentation, die den Aufbau des Formats und etwaige zulässige Datentypen in natürlicher Sprache erläutert.

5.2.3 Langzeitarchivierung und Langzeitarchivierungsstrategie

In jeder Forschungseinrichtung erhöht sich die Menge wissenschaftlicher Publikationen, Forschungsdaten und anderer digital gespeicherten Daten dadurch, dass viele ursprünglich analog vorliegende Daten digitalisiert werden. Viele Organisationen und Institutionen wollen damit den Zugriff und die Nutzung dieser Informationen über Datennetze vereinfachen (Schwens & Liegmann 2004). Zusätzliche Aufwände entstehen durch die Verwaltung des Datenarchivs und die Bereitstellung der Inhalte, wobei verschiedenartige Zugriffsrechte und Berechtigungen definiert werden. Diese einzeln aufgelisteten Komponenten erfordern nicht nur personelle Kapazitäten und Kompetenzen für Aufbereitung und Wartung, sondern auch technische Ressourcen wie Speicherplatz und Rechenzeit, sowohl zum Zeitpunkt der Bereitstellung als auch die Pflege der Forschungsprimärdaten. Die Entwicklung einer organisatorischen und technischen Infrastruktur zur systematischen Sicherung, Archivierung und Bereitstellung von Forschungsprimärdaten ist eine komplexere Aufgabe. Dabei sollten sowohl organisatorische Leitlinien als auch deren technische Realisierung sowie das Datenmanagement einbezogen und strategisch aufgebaut werden. Zu den großen Herausforderungen der Langzeitarchivierung gehört, dass sich nicht nur die Formate der Daten ändern, sondern auch die digitalen Datenträger, auf denen sie gespeichert sind. Diese können möglicherweise in relativ kurzer Zeit aufgrund des technischen Fortschritts nicht mehr gelesen werden; eine begrenzte physische Haltbarkeit der Datenträger oder die Alterung der Trägerformate kommen erschwerend hinzu. Im Allgemeinen wird zwischen drei Strategien der digitalen Langzeitarchivierung unterschieden, welche die langfristige Haltbarkeit der Daten garantieren können (Neuroth et al. 2010; Görl et al. 2011).

Bei der *Datenträgermigration* werden die Daten von einem Träger vor drohendem Datenträgerausfall auf einen anderen Datenträger kopiert, wobei es sich entweder um den gleichen oder ein anderen Datenträgertyp handeln kann, z.B. von Festplatte auf CD, von DVD auf Band etc. Diese Strategie ist schnell einsetzbar und leicht realisierbar. Sie erfüllt ihren Zweck, Daten von einem Datenträgerausfall zu bewahren, löst aber als solche das Problem nicht dauerhaft; eine Datenträgermigration muss in gewissen Zeitabständen immer wieder und rechtzeitig durchgeführt werden und vermag das Verhalten von Datenformaten nicht aufzuhalten.

An diesem Punkt setzt vielmehr die *Daten- oder Formatmigration* an. Dieser Ansatz zielt darauf ab, Formate in ein möglichst standardisiertes und offenes Format umzuwandeln, bevor ggf. proprietäre Vorgängerlösungen veralten. Der Vorteil dieser Maßnahme besteht darin, die Daten in einem neuen zeitgemäßen Format repräsentieren zu können, welches, sofern geeignet, das Potenzial einer langfristigen Lesbarkeit ungeachtet des technologischen

[38] Für eine ausführlichere Darstellung der Zusammenhänge von offenen Standards, Interoperabilität und Wettbewerb sei auf (Weston & Kretschmer 2012) verwiesen.

Wandels bietet. Bei der Formatmigration besteht jedoch stets das Risiko des Informationsverlustes. So kann es vorkommen, dass sich das äußere Erscheinungsbild der Daten ändert oder im schlimmsten Fall einige Teile verlorengehen. Aufgrund der Tatsache, dass ein Format ausgewählt werden muss, das nicht nur aktuell, sondern auch allgemein als Standard anerkannt ist und noch wahrscheinlich noch möglichst lange Zeit gewartet und benutzt zu werden verspricht, ist eine Formatmigration ein recht anspruchsvoller Prozess. Da diese Vorgehensweise ihrerseits nicht vor dem Ausfall von Datenträgern schützt, muss sie den zuvor beschriebenen Ansatz der Datenmigration flankieren, um vollwertigen Schutz zu bieten.

Als dritte Möglichkeit bietet sich die *Emulation digitaler Objekte* an, mithilfe derer die originale Umgebung der archivierten digitalen Objekte in einer aktuellen Software-Hardware-Umgebung nachbildet wird. Die Vorteile der Emulation bestehen sicherlich darin, dass sie Eingriffe in die digitalen Daten selbst erspart. Die Daten bleiben unverändert und können weitgehend originalgetreu genutzt werden. So kann auch ein Informationsverlust vermieden werden, wie er bei der Formatmigration bisweilen auftritt. Allerdings stellt die Realisierung auch dieser Maßnahme sehr hohe Ansprüche an Forschungseinrichtungen. Abhängig von der Komplexität der Objekte und Systeme einerseits und der Heterogenität der nachzubildenden Formate andererseits, steigt die Zahl der zu entwickelnden Emulatoren und mithin Aufwand und Kosten für ihre Implementierung.

Die Anwendung des Migrations- oder Emulationsverfahrens ist dann notwendig, wenn andere Maßnahmen für eine weitere Langzeitarchivierung der digitalen Objekte bereits ausgeschöpft sind. Um dies zu vermeiden oder zumindest zu verzögern fördern verschiedene Initiativen und Organisationen (nestor, DIN NABD 15 - NA[39]) die Verwendung langzeitstabiler Datenformate und offener Standards (vgl. Abschnitt 5.2.2 sowie (Schwens & Liegmann 2004; Neuroth et al. 2010)).

6. Begleitforschung zu Forschungsinfrastrukturen

Forschungsinfrastrukturen können nur so gut sein wie das Maß an Wissenschaftlichkeit, das sie abbilden, ja wie der Grad ihrer Verankerung in der wissenschaftlichen Aktualität.

„Viele Forschungsinfrastrukturen vereinen Elemente der Grundversorgung und der thematisch fokussierten und häufig projektförmigen Forschung. Die Übergänge sind fließend. Auch mag eine Infrastruktur als spezielles Forschungsprojekt mit vergleichsweise wenigen Nutzerinnen und Nutzern beginnen und sich dann zu einer Infrastruktur für weite Kreise einer Fachgemeinschaft ausweiten. Dies zeigt ein Blick auf die Genese der erfolgreichen Forschungsinfrastrukturen in den Geistes- und Sozialwissenschaften. Zahlreiche dieser Einrichtungen haben als zeitlich befristete Forschungsprojekte begonnen und sich dann im Laufe der Jahre zu Einrichtungen der Grundversorgung in Forschung und Lehre entwickelt, die nicht nur einen bestehenden Bedarf abdecken, sondern neuen Bedarf schaffen, indem sie durch me-

[39] Informationstechnik und Anwendungen (NIA) im DIN - Schriftgutverwaltung und Langzeitverfügbarkeit digitaler Informationsobjekte, http://www.din.de

thodische Innovationen und durch die Generierung interessanter Fragestellungen zunehmend weitere Kreise von Nutzerinnen und Nutzern einbeziehen.

Die Grundausrichtung einer Forschungsinfrastruktur lässt sich somit nicht von vornherein ein für alle Mal festschreiben. Sie muss sich im Laufe der Zeit ihre Adaptionsfähigkeit an eine dynamische Forschungslandschaft erhalten." (Wissenschaftsrat 2011a, S. 23f.) heißt es aus diesem Grund auch in den Empfehlungen des Wissenschaftsrates zu Forschungsinfrastrukturen in den Geistes- und Sozialwissenschaften. Umgesetzt wurde diese Forderung durch die Initiierung einer bundesweiten Roadmap für Forschungsinfrastrukturen, an deren Rahmenprogramm sich der Aufbau von Forschungsinfrastrukturen orientieren und die Lösungsansätze der Trägereinrichtungen bewegen sollten. Dieser angestoßene Prozess soll dazu dienen, *„forschungspolitische Entscheidungen hinsichtlich der Zuordnung von beschränkten Ressourcen für Forschungsinfrastrukturen vorzubereiten und zu unterstützen. Alle Vorhaben in den verschiedensten Wissenschaftsgebieten und über alle potenziellen Trägerorganisationen hinweg werden in die Überlegungen zu Bedarf an Forschungsinfrastrukturvorhaben, deren Zielsetzung und Qualität sowie deren Kosten im Aufbau und Betrieb einbezogen*" (Bundesministeriums für Bildung und Forschung 2013, S. 3).

Ziel ist hierbei eine dauerhafte Verankerung von Serviceeinheiten in ihren Trägerinstitutionen. Eine kontinuierliche Begleitforschung sichert hierbei die Qualität und Relevanz von Forschungsinfrastrukturen in und für die zu unterstützenden Wissenschaftler. Forschungsinfrastrukturen müssen hierbei mit der Entwicklung aller ausschlaggebenden Rahmenbedingungen Schritt halten können. Forschungsinfrastrukturforschung adressiert daher die Schlüsselthemen aus dem Dunstkreis des wissenschaftlichen Dienstleistungssektors, um die Dienste nicht nur den Fragestellungen der Wissenschaft anzupassen, sondern auch technologisch, administrativ und juristisch nachhaltig bestehen zu können. Die Landschaft dieser Begleitforschung stellt sich in seiner Themenvielfalt demnach ausgesprochen heterogen dar. Nicht zuletzt empfiehlt der Wissenschaftsrat daher Kooperation von Fachwissenschaftlern mit Spezialisten serviceorientierter Sachgebiete (Wissenschaftsrat 2012, S. 66). Die Aufgabenfelder erstrecken sich hierbei über Rahmenbedingungen, Finanzierung, Planung, Organisation und Nutzung des Serviceangebots (Wissenschaftsrat 2011b, S. 30ff.), was analog zur Erarbeitung einer Systematik der Bewertungskriterien zur Förderung von Forschungsinfrastrukturen im Rahmen der *Roadmap* geführt hat, die künftig über zwei Schienen erfolgen soll (Bundesministeriums für Bildung und Forschung 2013, S. 4). Der wissenschaftsgeleitete Bewertungsprozess bewertet Forschungsinfrastrukturen per Peer-Review (vgl. Wissenschaftsrat 2011a, S. 24) hinsichtlich ihres wissenschaftlichen Potenzials (Wechselspiel aus Serviceangebot und Forschungsstand), ihres Nutzungspotenzials (offene Architektur, Nutzerkreis, Qualitätsanspruch), ihrer Umsetzbarkeit (technischer Stand) sowie ihrer Bedeutung für den Wissenschaftsstandort Deutschland, kurz auf Kriterien der technischen wie fachwissenschaftlichen Nachhaltigkeit, abgeklopft. Der wirtschaftliche Bewertungsprozess untersucht hingegen die Konzeption der finanziellen und organisatorischen Nachhaltigkeit einer Forschungsinfrastruktur. Nicht alle diese Kriterien spielen beim Aufbau von Forschungsinfrastrukturen in ihren Trägerorganisationen eine Rolle bzw. sind vor Ort zu lösen, sondern sprechen vornehmlich Förderer und Zuwendungsgeber an, sodass hier nur ausgewählte Punkte herausgegriffen werden sollen.

Als nationales Anliegen werden Forschungsinfrastrukturen zumeist aus öffentlicher Hand *finanziert*. Jedes Finanzierungsmodell hat somit die langfristige Funktionalität von Forschungsinfrastrukturen mitzudenken. Eine projektbasierte Förderung kann somit nicht die Lösung eines Dauerbetriebes sein. Vielmehr sind Forschungsinfrastrukturen so in ihren Trägerinstitutionen zu verankern, dass sie dauerhaft auf Akzeptanz der Forscher stoßen, indem ihr wissenschaftlicher Mehrwert gewahrt bleibt.

Fragen der *organisatorischen* Konzeption zielen insbesondere auf einen dauerhaften Bestand der Forschungsinfrastruktur vor dem Hintergrund ihrer finanziellen Rahmenbedingungen ab. Die Einbindung starker Trägerorganisationen und möglichst breiter Nutzerschichten ist die Grundlage eines jeden Organisationsmodells.

Bei der Planung von Forschungsinfrastrukturen hinsichtlich ihrer *Nutzung und Nutzbarkeit* kommt es vornehmlich auf ein koordiniertes Vorgehen in der Weiterentwicklung der Angebote an. Nicht nur müssen Synergien gerade mit Blick auf die beschränkten Möglichkeiten kleiner Fächer genutzt werden, um Redundanzen oder gar ein Zurückbleiben einzelner Disziplinen zu vermeiden, sondern auch proprietäre und somit technisch nicht nachhaltige Insellösungen durch die Adressierung zentraler Themen wie Standardisierung, Forschungs- und Metadatenmanagement, Langzeitarchivierung und anderer innovativer Entwicklungen umgangen werden. Um die Relevanz von Forschungsinfrastrukturen innerhalb der Nutzerschaft kontinuierlich zu erhalten, sind sie nicht nur organisatorisch nachhaltig aufzustellen, sondern auch technisch so zu gestalten, dass sie den (dynamischen) Bedarfen der Wissenschaft antworten. Außerdem ist die Thematik in den Curricula der akademischen Lehre zu verankern, um den wissenschaftlichen Nachwuchs in der Kenntnis um die Notwenigkeit von sowie im Umgang mit Komponenten von Forschungsinfrastrukturen entsprechend zu qualifizieren.

Außeruniversitäre Forschungsinstitute sind demnach nicht nur an der Bereithaltung der Forschungsinfrastrukturen interessiert, sondern aktiv daran beteiligt in diesem Bereich Forschung zu ihrer Weiterentwicklung betreiben. Als Beispiele seien an dieser Stelle lediglich die unter der Fördermaßnahme Wissenschaftliche Literatur- und Informationssysteme (LIS)[40] der Deutschen Forschungsgemeinschaft laufenden Projekte herausgegriffen. Ziel dieser Maßnahme ist der Aufbau von bedarfsgerechten Informationsinfrastrukturen in universitären und außeruniversitären Bereich. Hierbei werden Themenschwerpunkte wie Informationsdienste, überregionale Lizenzierung, Digitalisierung, digitale Editionen, Open Access, Virtuelle Forschungsumgebungen, Nachhaltigkeit und Forschungsdatenmanagement gesetzt.

Innerhalb der Ausschreibung „Informationsinfrastrukturen für Forschungsdaten“[41] aus der Fördermaßnahme LIS finden sich zahlreiche Projekte aus den Geistes- und Sozialwissenschaften. So wird mit dem Projekt „Integration von Forschungsdaten und Literatur in den Sozialwissenschaften“ (InFoLiS)[42] am GESIS Leibniz-Institut für Sozialwissenschaften ein Vorschlag für eine nachhaltige Lösung für die Interaktion unterschiedlichster Daten(-formate)

[40] http://www.dfg.de/foerderung/programme/infrastruktur/lis/index.html
[41] http://www.dfg.de/download/pdf/foerderung/programme/lis/projekte_forschungsdaten.pdf
[42] http://www.gesis.org/forschung/drittmittelprojekte/projektuebersicht-drittmittel/infolis/

gesellschaftsbezogener Forschung erarbeitet. Die Altertumswissenschaften stehen hingegen mit einer hohen Diversität an Forschungsdaten, -formaten, -verfahren und Metadaten vor großen Herausforderungen, ihre empirische Datengrundlage weiterhin nutzbar zu erhalten, die das Deutsche Archäologische Institut (DAI) mit dem LIS-Projekt „Entwicklung eines Kompetenzzentrums für altertumswissenschaftliche Forschungsdaten“[43] anzugehen versucht. Ähnliche Vorhaben der jeweilig disziplinspezifisch konzipierten Optimierung und Verstetigung ihres Forschungsdatenmanagements wurden in dieser Linie u.A. an der Universität Hamburg mit der Einrichtung eines Forschungsschwerpunkts „Mehrsprachigkeit und gesprochene Sprache“,[44] an der Berlin-Brandenburgischen Akademie der Wissenschaften (BBAW) mit dem Projekt „Wissensspeicher – Daten geisteswissenschaftlicher Grundlagenforschung“,[45] an der Universität Duisburg-Essen mit dem Aufbau eines Zentrums für Record-Linkage[46] oder dem IDS mit seinem Zentrum für germanistische Forschungsprimärdaten eingeleitet.

Andere Vorhaben jenseits dieser DFG-Linie wie die Projekte CLARIN-D,[47] das im europäischen Rahmen einen deutschlandweiten Zentrenverbund mit Expertise in Service- und Infrastrukturlösungen etabliert, die Virtuelle Forschungsumgebung TextGrid[48] oder das Netzwerk DARIAH[49] widmen sich neben technischen Themenschwerpunkten auch Fragestellungen der nachhaltigen Finanzierung, stabilen Organisation, der Nutzerbetreuung und des Community-Building sowie juristischen Aspekten des Umgangs Forschungsdaten.

Die Vielfalt all dieser Vorhaben zeigt die Komplexität der beim Aufbau von Forschungsinfrastrukturen zu bedenkenden Aspekte und deren Lösungen auf. Mit der einmaligen technischen Installation der Basisgeräte und –dienste – eine Aufgabe, die allein für sich genommen Forschungseinrichtungen bereits vor große Herausforderungen stellt – ist es daher in diesem Zusammenhang nicht getan. Es empfiehlt sich vielmehr dringend, die stetige Adaption dieser Anfangsausstattung, wie sie angesichts hoher technischer, sozialer, finanzieller und wissenschaftlicher Dynamik unverzichtbar ist, mit bedarfsorientierten Forschungsprojekten unterschiedlichster thematischer Ausrichtung zu begleiten. Dass einzelne Akteure hierbei stets den Kontakt und die Anknüpfung an bereits bestehende Lösungsversuche und die Nutzung von Verbundlösungen suchen sollten, versteht sich von selbst.

7. Interviews

Im Rahmen zahlreicher Projekte hat das Institut für Deutsche Sprache (IDS) eine Reihe von Fertigkeiten rund um den Arbeitsschwerpunkt Forschungsinfrastrukturen akkumuliert. Hierunter fallen Forschungen zu sowie die Etablierung und Straffung von Infrastrukturen und des Managements analoger und digitaler Forschungsdaten mit vereinheitlichten Standards und nachhaltigen Organisationsmodellen. Hierbei hat sich gezeigt, dass der Aufbau einer kombi-

[43] http://www.ianus-fdz.de/projects/zentr-dig-arch/wiki/Wiki?version=1
[44] http://virt-fedora.multilingua.uni-hamburg.de/drupal/node/35
[45] http://www.bbaw.de/telota/ressourcen/digitaler-wissensspeicher
[46] https://www.uni-due.de/gesellschaftswissenschaften/profilschwerpunkt/record_linkages.shtml
[47] http://de.clarin.eu/de/home/projektueberblick.html
[48] http://www.textgrid.de/ueber-textgrid/projekt/
[49] https://de.dariah.eu/about

nierten Basis aus Infrastrukturforschung und Forschungsdienstleistung gleichwohl anspruchsvoll wie nutzbringend sein kann. Im Wissen um die Vorteile, die sich aus solchen Maßnahmen ergeben, strebt manches Forschungsinstitut derzeit ebenso nach einer Anpassung seiner Forschungsinfrastrukturen an die Anforderungen modernen Datenmanagements. Obschon die Ansätze und Voraussetzungen in den einzelnen Instituten von Disziplin zu Disziplin hinsichtlich Forschungsschwerpunkt oder Methodik bisweilen recht stark variieren, zeigen sich dennoch Anknüpfungspunkte für ein konzentriertes Vorgehen. Die Nutzung der bereits vorhandenen Erfahrungen und das Profitieren von Synergien aus einem abgestimmten Handeln könnten somit nicht nur in gegenseitigem Erkenntnisgewinn resultieren, sondern böten vielmehr auch in einem gemeinsamen Nenner einen Ausgangspunkt für die Entwicklung moderner und kompatibler Forschungsinfrastrukturen in den einzelnen Häusern. Gerade kleinere Einrichtungen, die nicht minder auf eine zeitgemäße Ausstattung Wert legen, viele Anschaffungen aber nur im Verbund oder in Zusammenarbeit mit externen Partnern stemmen können, zählen zu den besonderen Profiteuren eines vernetzten Handelns.

Um sich zu Beginn solcher Überlegungen einen Eindruck zu verschaffen, wie sich der Stand der Arbeiten im Aufbau von Forschungsinfrastrukturen derzeit bei außeruniversitären Forschungseinrichtungen darstellt sowie die verschiedenen Ansätze, Möglichkeiten und Meinungen auf ihrem bisherigen Weg auszuloten, hat das IDS im Auftrag des BMBF eine qualitative Interviewserie bei solchen Institutionen durchgeführt. Hierfür wurden zwölf Häuser aufgesucht, von denen sich vier trotz eines deutlichen Forschungsanteils – nicht zuletzt handelt es sich um infrastrukturbezogene Forschung als solche – als dedizierte Forschungsinfrastrukturen verstehen. Eine der befragten Einrichtungen bildet zwar als Stiftung den Rahmen für zahlreiche Kulturdenkmäler und macht sie der Öffentlichkeit zugänglich, doch hat der Forschungsanteil in diesem Fall ein solches Gewicht inne, dass sie unter den klassischen Forschungsinstituten subsumiert wurden.

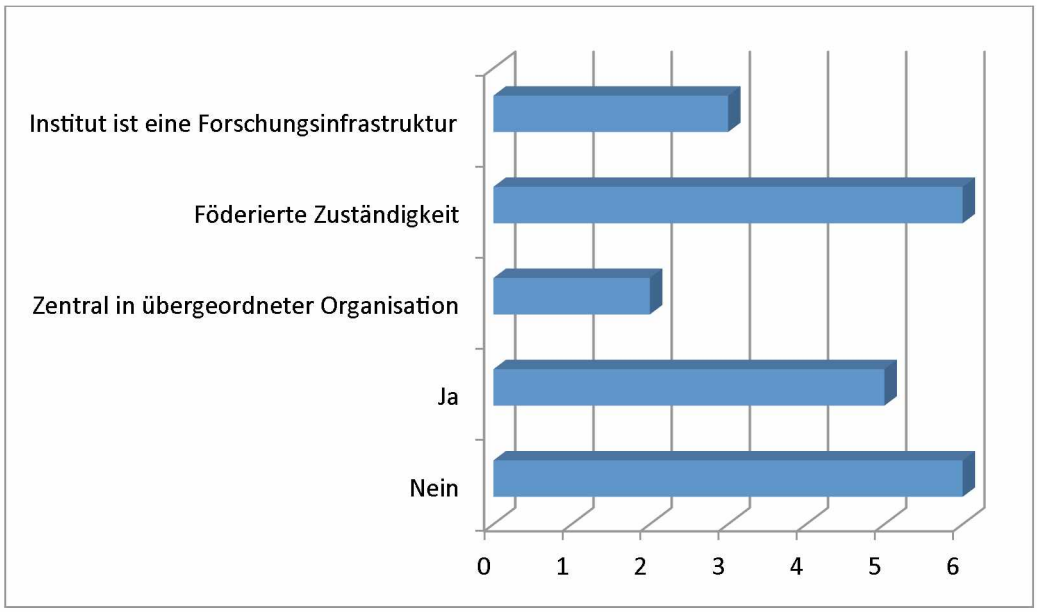
Für die Befragung wurde ein standardisierter Leitfaden erarbeiten, der aus den oben beschriebenen Aspekten die Kernkomponenten moderner Forschungsinfrastrukturen herausfiltert und thematisiert. So wurden die Fragen in Rubriken zu Organisation und Zuständigkeiten, Datenmanagement, Hardware und Software, Umweltschutz sowie Recht zusammengefasst. Alle Befragten haben der Verarbeitung ihrer Antworten zu einer zusammenfassenden und vergleichenden Analyse zugestimmt, wobei keine personenbezogenen Daten erhoben und die Interviews an Schlüsselstellen anonymisiert wurden. Desgleichen werden die angesprochenen Institutionen und ihre Vertreter in der folgenden Analyse nicht namentlich benannt. Die Beschaffenheit des Instituts für Deutsche Sprache selbst wurde zwar entlang dieses Leitfadens ebenso erfasst, die Ergebnisse wurden jedoch aus methodischen Gründen nicht in die Analyse aufgenommen. Aufgrund des qualitativen Charakters der Befragung und der starken methodischen und inhaltlichen Divergenz der angesprochenen Institute deckte der vorgegebene Befragungsstandard nicht in allen Fällen das Profil der Interviewten ab, sodass sich im Bedarfsfalle Anpassungen bei der Auswertung ergaben.

Der Verlauf der Interviews wurde während des Gesprächs aufgezeichnet und anschließend transkribiert. Für die qualitative Auswertung erfolgte eine Zuordnung sämtlicher Aussagen zu den Rubriken des Fragebogens, um in einem weiteren Schritt systematisiert und kategorisiert zu werden. Die Zahl der in die Auswertung einbezogenen Kategorien legte hierbei Wert

auf Vollständigkeit und nicht etwa auf Quantität und Übereinstimmung der jeweiligen Antworten. Im Sinne der Anonymisierungen erfolgte die graphische Darstellung dergestalt, dass sich keine Zuordnung von Antworten zu einzelnen Instituten vornehmen lässt, um eine Profilierung oder gar Identifizierung einzelner Interviewpartner zu erschweren.

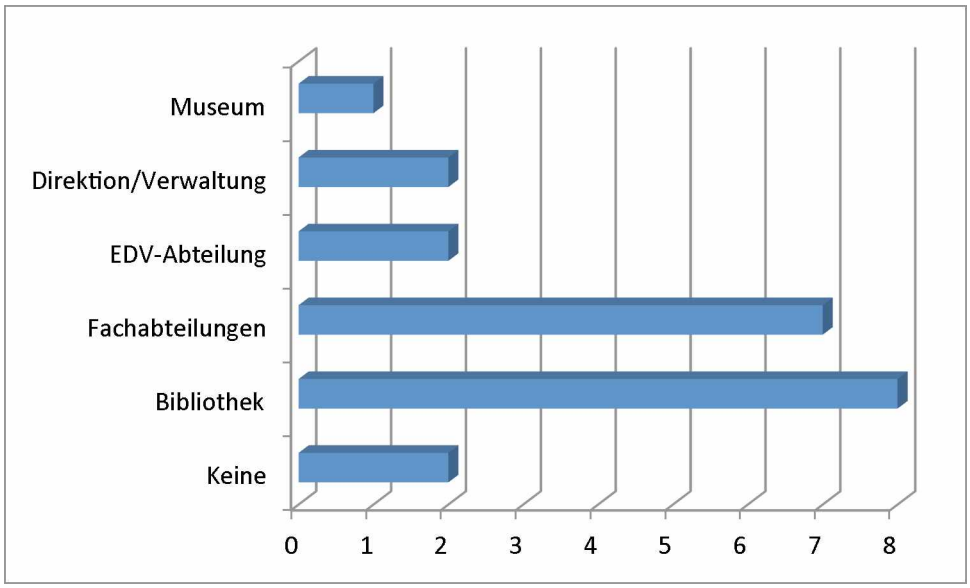
7.1 Organisatorische Aspekte

Frage 1: *Gibt es eine verantwortliche Organisationseinheit für Forschungsinfrastrukturen (z.B. eine/n CIO)?*



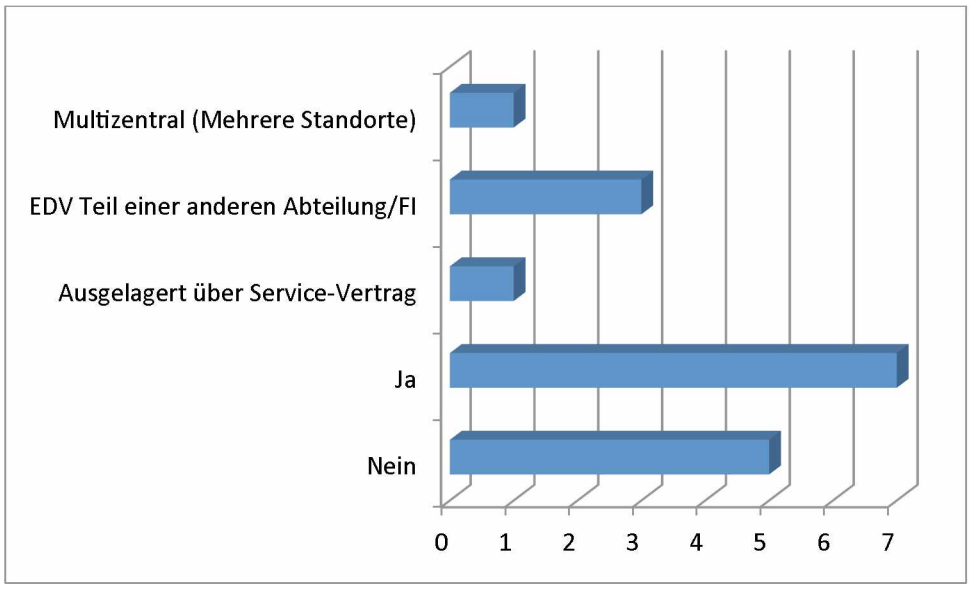
Im Rahmen der Frage nach der organisierten Verfasstheit von Forschungsinfrastrukturen innerhalb der befragten Institutionen differenzierten sich rasch die Unterschiede im Selbstverständnis von Bibliotheken, Museen oder Archiven und reinen Forschungsinstituten heraus: Erstere bezeichnen sich entweder selbst als vollwertige Forschungsinfrastruktur oder ordnen sich übergeordneten Infrastruktureinheiten zu. Im Falle der Forschungsinstitute stellt sich die Situation differenzierter dar. Fünf der neun Befragten weisen keine dediziert für Forschungsinfrastrukturen zuständige Organisationseinheiten wie etwa einen Chief Information Officer auf. Vielmehr verteilen sich infrastrukturelle Einheiten wie IT-Abteilung oder Institutsbibliothek entweder dezentral über die gesamte Einrichtung oder sind größeren (Fach-)Abteilungen zugeordnet. Gerade bei einer dezentralen Organisationsform überwiegen föderierte Zuständigkeiten, im Zuge derer sich einzelne Abteilungen die Organisation der Serviceabteilungen in den Häusern teilen – bisweilen werden diese Aufgaben durch Gremien koordiniert. Nur zwei der insgesamt zwölf Interviewpartner zeigen ein hohes Maß an Integration der Forschungsinfrastrukturen in einer zentralen Organisationseinheit. Dies gestaltet sich zumeist in Form einer spezialisierten Abteilung, die sich sowohl als Forschungs- als auch als Servicekomponente für die jeweilige disziplinspezifische Fragestellung versteht.

Frage 2: *Welche (weiteren) Organisationseinheiten sind mit Forschungsinfrastrukturen-Fragen befasst?*



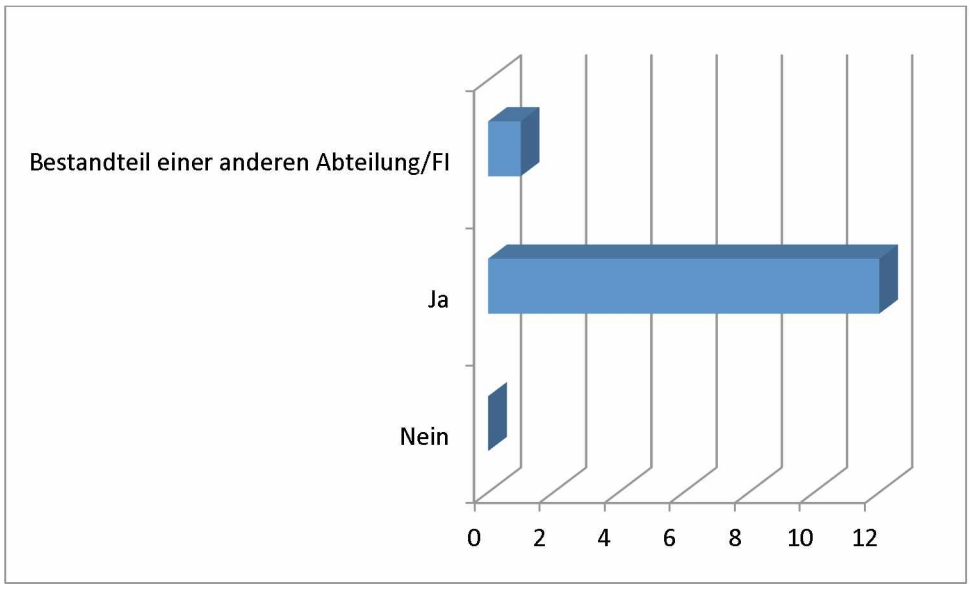
In Ihrem Verständnis von Forschungsinfrastrukturen nennt die überwiegende Mehrheit der Institute die Bibliothek als grundlegende Serviceeinrichtung ihres Hauses. Etwas weniger sehen sie disziplinspezifische Fachabteilungen in dieser Reihe. Bei den Instituten ist dieser Umstand jedoch unabhängig von der sonstigen organisatorischen Verfasstheit der Forschungsinfrastrukturen in ihren Häusern zu sehen. So antworteten die sechs befragten Institute, die Forschungsinfrastrukturen in Fachabteilungen ansiedeln, auf die vorherige Frage nach dem Grad der Zentralisierung je hälftig, ihre Dienstleister seien sehr integriert bzw. kaum zentralisiert oder föderiert verfasst. Darüber hinaus werden, soweit als selbstständige Einheit vorhanden, die EDV-Abteilung bzw. die Leitungsebene genannt. Infrastruktureinrichtungen verweisen hingegen wiederum entweder auf übergeordnete Strukturen bzw. – als eigenständige Infrastruktur – auf ihre thematisch spezialisierten Fachabteilungen oder Verwaltungsebenen, in einem Fall auf die Bibliothek.

Frage 3: *Gibt es in Ihrem Institut eine zentrale EDV-Einheit (z.B. ein Rechenzentrum, etc.)?*



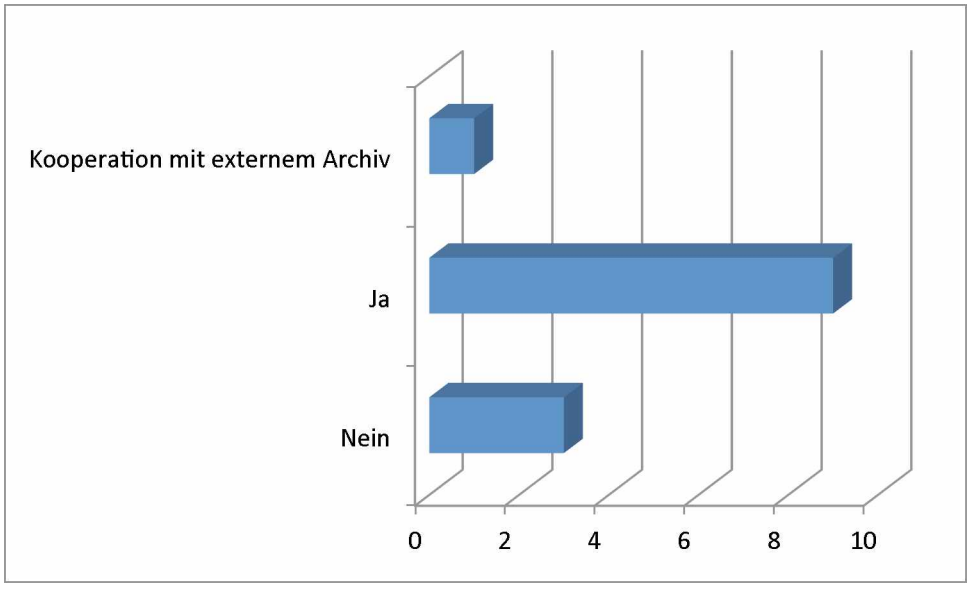
Über die Hälfte der befragten Einrichtungen verfügen über eine eigene IT-Abteilung (sieben von zwölf); auf fünf weitere trifft dies nicht zu. Von den Befragten, die diese Frage verneinten, ordnen drei die IT einer anderen Abteilung unter (zumeist der Verwaltung), einer lagert sie an einen externen Dienstleister aus – in diesem Falle die standortnahe Universität – und einer verweist analog zu den Fragen zuvor an eine übergeordnete Ebene. Lediglich ein Institut weist aufgrund seiner standortübergreifenden Organisation mehrere IT-Abteilungen auf.

Frage 4: *Gibt es in Ihrem Institut eine Bibliothek?*



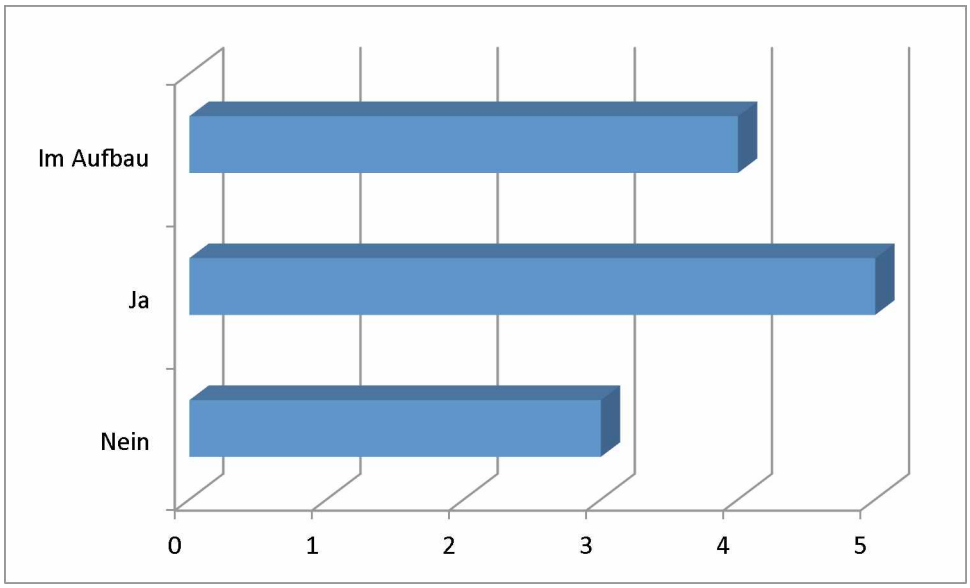
Wie sich zeigte, verfügt jede befragte Institution über eine eigene Bibliothek. Eine Einrichtung, die in Frage 2 die Bibliothek nicht den Forschungsinfrastrukturen zuordnete, weist nichtsdestotrotz eine solche Einheit auf. Nur eine Institution hat die Bibliothek in ihre übergeordnete Forschungsinfrastruktureinheit integriert.

Frage 5: *Gibt es in Ihrem Institut ein oder mehrere Archiv(e)?*



Auf die Frage nach der Existenz von Archiven antworteten nur drei der Befragten nicht positiv. Eines dieser Institute lagert seine zu archivierenden Unterlagen extern aus. Da die meisten Archive der Ablage von administrativen Dokumenten dienen, wird diese Komponente meist nicht unter die Forschungsinfrastrukturen gezählt.

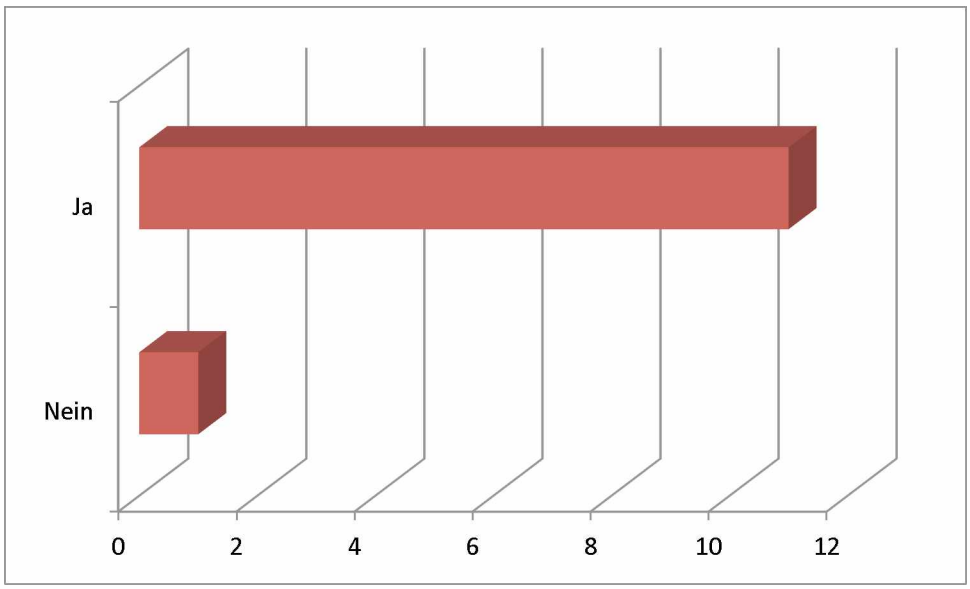
Frage 6: *Gibt es in Ihrem Institut (zentrale oder lokale) Repositorien?*



Sehr viel heterogener stellt sich die Situation hinsichtlich der Vorhaltung von zentralen oder lokalen Datenrepositorien dar. Fünf der zwölf befragten Einrichtungen beantworteten diese Frage positiv, während vier Interviewpartner dies verneinten. Je nach Größe und disziplinärer Fragestellung haben viele Institutionen die Vorteile eines solchen Datenspeichers erkannt und sind derzeit im Aufbau von Repositorien begriffen.

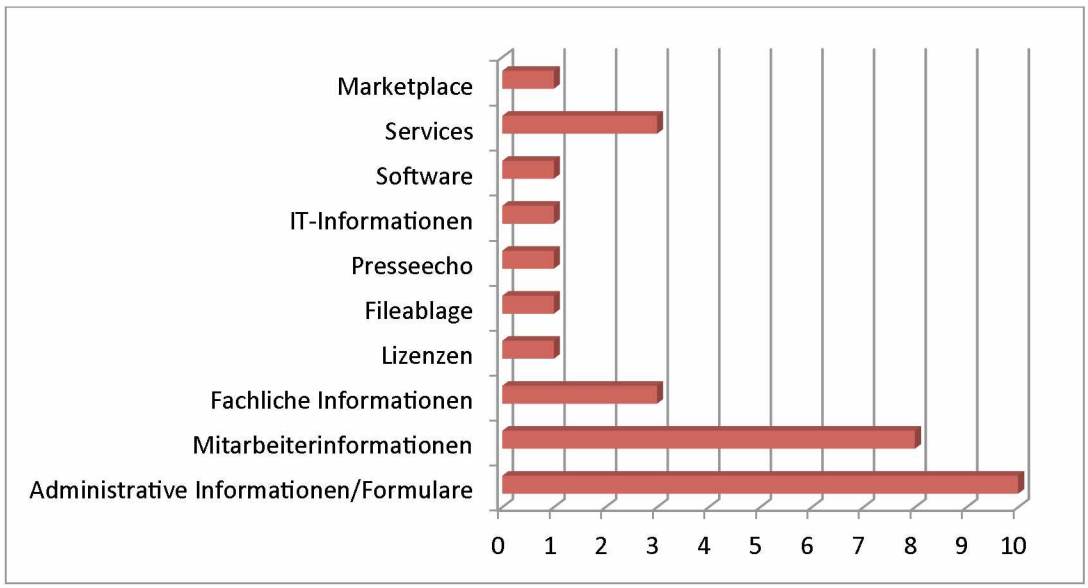
7.2 Internet-Intranet

Frage 7: *Gibt es Ihrem Haus ein Intranet?*



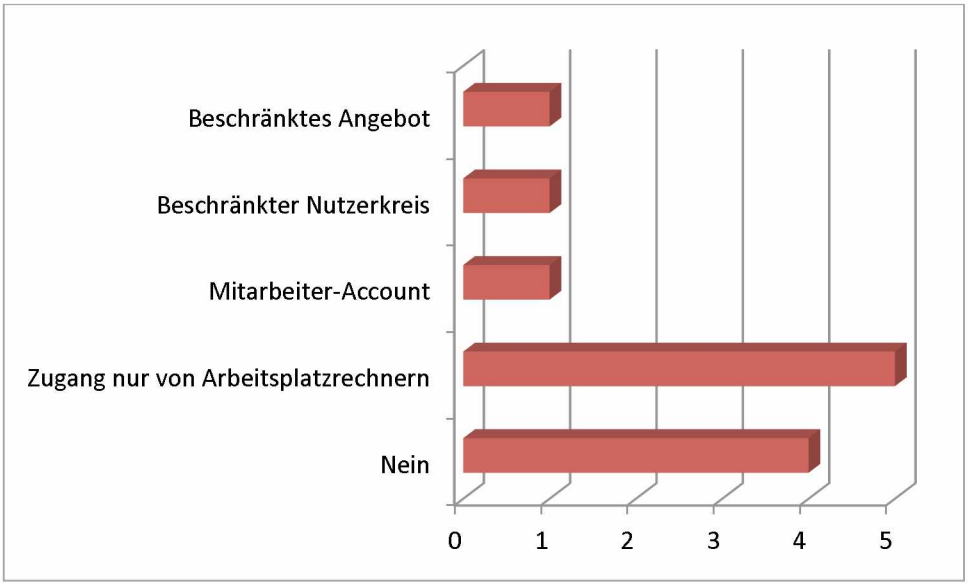
Unter den (netzbasierten) Diensten wird das Intranet im Regelfall zur Verfügung gestellt. Wo dies nicht der Fall ist, ist die geringe Institutsgröße als Ursache anzuführen.

Frage 8: *Welche Angebote werden im Intranet bereitgehalten?*



Das Intranet dient in allen Institutionen der Bereitstellung administrativer Formulare und Informationsmaterialen. Eng damit verbunden, aber weniger häufig, können Mitarbeiter der befragten Einrichtungen dem Intranet Informationen über Arbeitskollegen (Arbeitsschwerpunkt, Publikationen, Lebenslauf, personenbezogene Daten) entnehmen. Wesentlicher seltener werden fachliche Informationen, also Forschungsdaten, oder spezielle Dienste angeboten. Andere Inhalte wie Vorlagen für Lizenzen, Fileablage, Presseecho usw. stellen Einzelfälle dar.

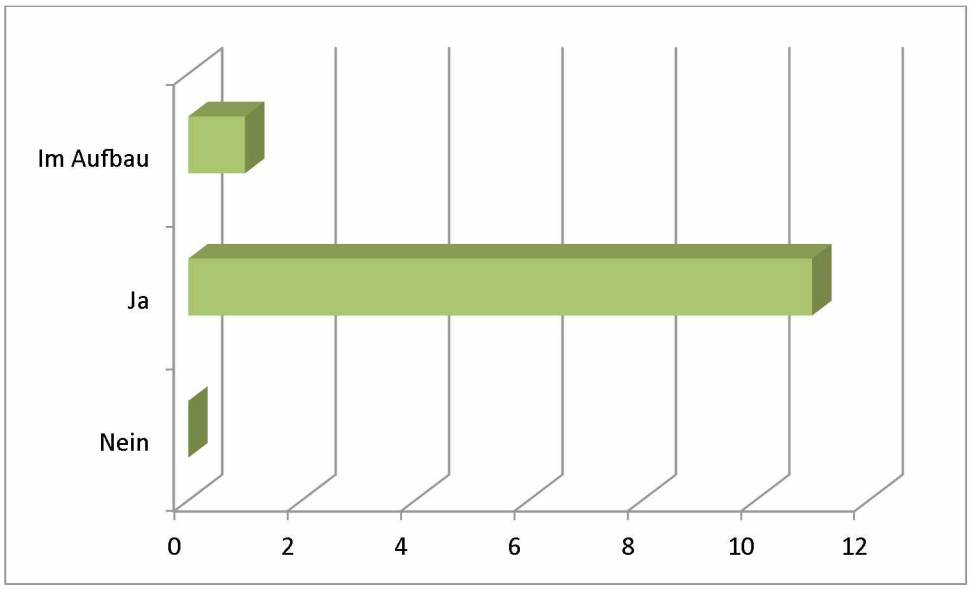
Frage 9: *Gibt es technische, organisatorische und/oder juristische Einschränkungen bei der Nutzung von Intranet-Angeboten und anderen internen Diensten von externen Arbeitsplätzen?*



Um den Intranet-Zugang, insbesondere zu personenbezogenen Daten, ggf. zu reglementieren, besteht die Möglichkeit, die Angebote gerade auch von externen Zugängen (VPN etc.) sowohl über rechtliche als auch, in der Folge, über technische Schranken zu verwehren oder einzugrenzen. Vier der zwölf Befragten beschränken den Zugang auf keine der genannten Weisen. Die häufigste Art der Reglementierung stellt die Freischaltung des Intranets ausschließlich über Arbeitsplatzrechner dar. Als weitere Methoden werden die Vergabe von dedizierten Accounts, die Einschränkung des Kreises der Zugangsberechtigten (Instituts-, Abteilungs-, Projektleitung) sowie die Einschränkung des Angebots für bestimmte Nutzerkreise (z.B. Hausgäste) genannt.

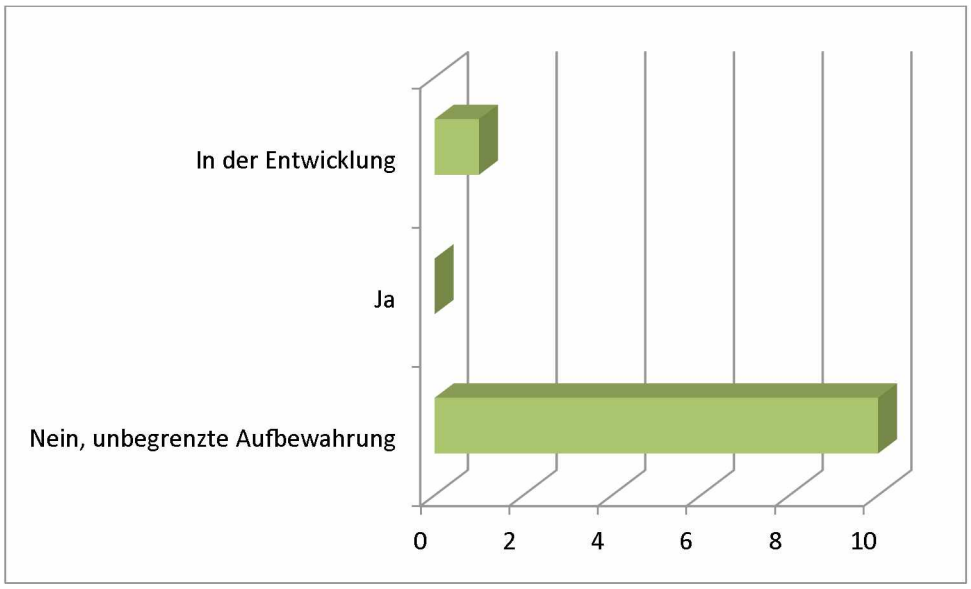
7.3 Datenbestand

Frage 10: *Verwalten (und pflegen) Sie in Ihrem Institut Forschungsprimärdaten?*



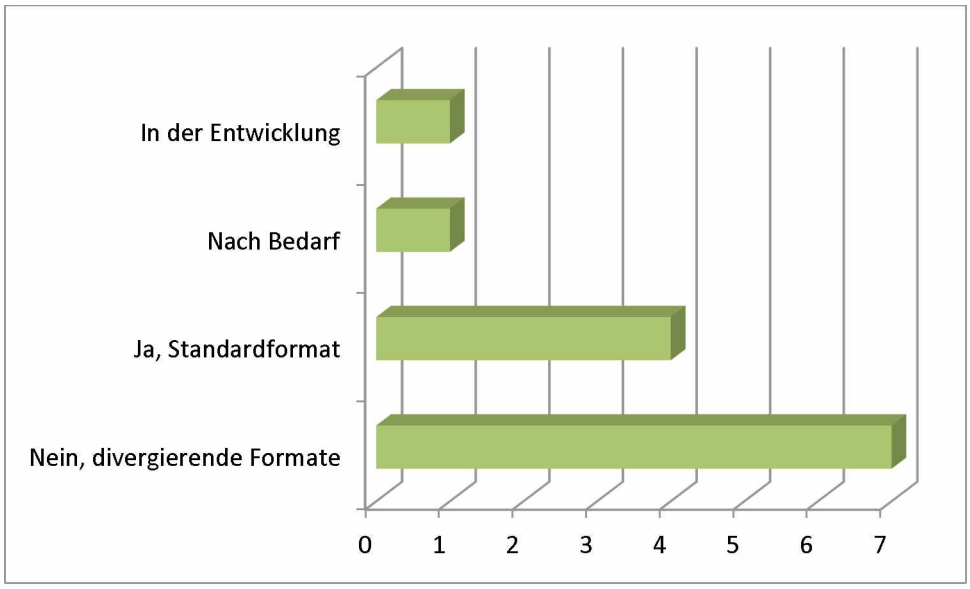
Die zunehmende Bedeutung von (digitalen) Forschungsprimärdaten als empirische Grundlage wird anhand der Tatsache deutlich, dass sämtliche befragten Institutionen bereits jetzt einen Bestand unterschiedlich verfasster und definierter Primärdaten vorhalten, pflegen, oder zumindest gerade aufbauen. Es wurde im Rahmen der Befragung deutlich, dass das Verständnis über Art und Funktion dieser empirischen Basis von Institut zu Institut stark variiert und insbesondere die Grenzen zwischen unbearbeiteten Rohdaten und aufbereiteten Primärdaten zusehends verläuft, weswegen diese Trennung bei der Analyse nicht vorgenommen wird.

Frage 11: *Gibt es festgelegte Aufbewahrungsfristen (ggf. Kurationszeitpläne)?*



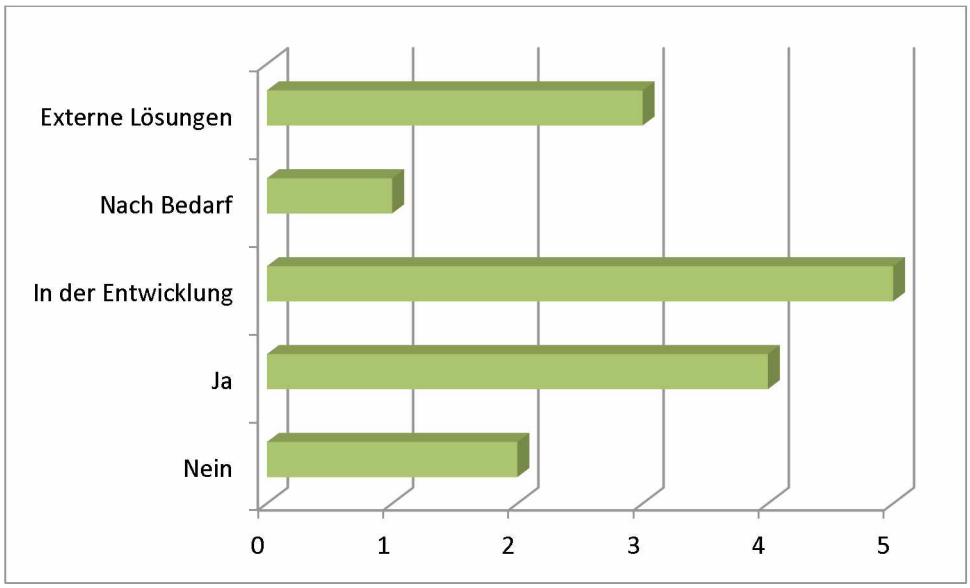
Forschungsprimärdaten bilden den Ausgangspunkt für alle empirisch geführten Beweisketten. Folglich ist die Verifizierbarkeit von Forschungsergebnissen primär abhängig von der bloßen Weiterexistenz (älterer) Forschungsprimärdaten. Zu diesem Zweck kann die Priorisierung von Beständen nach Aufbewahrungsdauer ein Kernbestandteil von Forschungsdatenmanagement sein. Vor diesem Hintergrund begründete die große Mehrheit der Befragten ihre verneinende Antwort auf die Frage nach der Limitierung von Aufbewahrungszeiträumen mit der Absicht, ihren Datenbestand aus Gründen der Nachvollziehbarkeit und im Sinne des wissenschaftlichen Institutsauftrages dauerhaft vorhalten zu wollen. Lediglich ein Institut beabsichtigt perspektivisch eine Abstufung der Kuration.

Frage 12: *Gibt es einen Konvertierungsplan bzw. einen Plan zur Migration zu neuen oder vereinheitlichten Speicherformaten?*



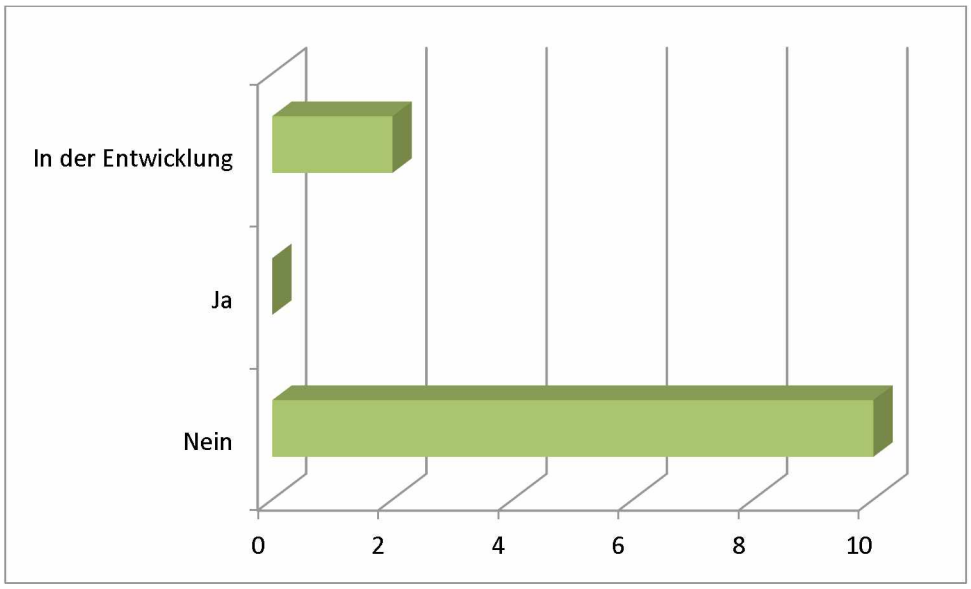
Die Datenablage in divergierenden, nicht nachhaltig gepflegten, proprietären oder schlichtweg nicht standardisierten Datenformaten ist in der Datenarchivierung einer der Hauptgründe für permanenten Datenverlust. Die Frage nach dem Stand der Vereinheitlichung der Formate beantwortete eine Mehrheit der Befragten abschlägig: die Formate der vorgehaltenen Daten divergieren innerhalb der einzelnen Einrichtungen teilweise stark. Bisweilen reichen die Kapazitäten der Befragten kaum aus bzw. werden vollständig dafür in Anspruch genommen, nichtmehr lesbare Formate in für zeitgemäße Anwendungen interpretierbare Formate zu konvertieren. Vier der Interviewten, davon zwei der Infrastruktureinrichtungen, benutzen Standardformate bzw. befinden sich in der Standardisierungsphase. Lediglich in einem Institut wird diese Frage je nach Bedarf der Datennutzer flexibel gehandhabt.

Frage 13: *Haben oder planen Sie eine Langzeitarchivierungsstrategie?*



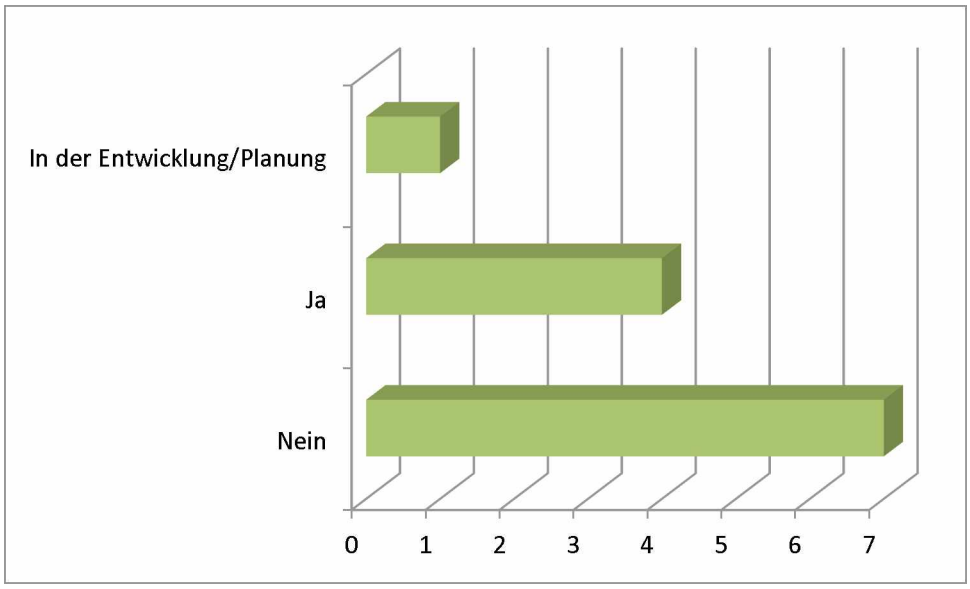
Langzeitarchivierung von Daten ist oft ein wesentliches Element modernen Forschungsdatenmanagements. Trotz des teils recht hohen Aufwandes der Umsetzung einer Langzeitarchivierungsstrategie und der momentan noch vollkommen ungeklärten Kostensituation für eine dauerhafte Aufbewahrung haben die Forschungsinstitutionen die Dringlichkeit dieser Aufgabe erkannt. So antworteten vier Institute, sie haben bereits eine solche Strategie umgesetzt, und weitere fünf entwickeln derzeit langfristige Lösungsansätze. Ein der hohen Komplexität der Umsetzung geschuldete Drang zu Kooperation hat drei der Institute mit eigener LZA-Strategie dazu veranlasst, auf externe Lösungen zurückzugreifen. Lediglich zwei Institute sehen aufgrund geringer Größe bzw. geringen Bedarfs hinsichtlich andersgearteter Forschungsdaten keinen Handlungsbedarf, während eines zwar Lösungen bereithält, sie aber nicht verbindlich und nur im Bedarfsfall für seine Mitarbeiter umgesetzt.

Frage 14: *Haben Sie ein Langzeitarchivierungssystem implementiert?*



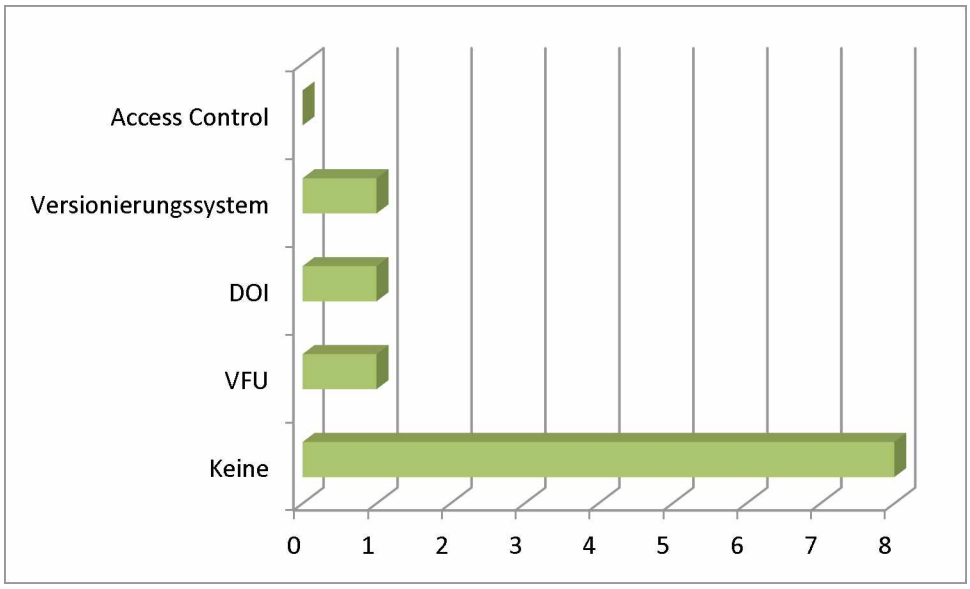
Bei den in den einzelnen Einrichtungen vorherrschenden Ansätzen für die Langzeitarchivierung fällt auf, dass fast alle institutsinterne, proprietäre Lösungen für ihre Datenhaltung gewählt haben. Lediglich eine Infrastruktureinrichtung greift auf ein umfassendes, servergestütztes LZA-System zurück. Bei dreien ist diese Situation auf die Auslagerung dieses Dienstes zu externen Partnern zurückzuführen.

Frage 15: *Verfolgen Sie im Rahmen der Replizierbarkeit von Forschungsergebnissen Archivlösungen?*



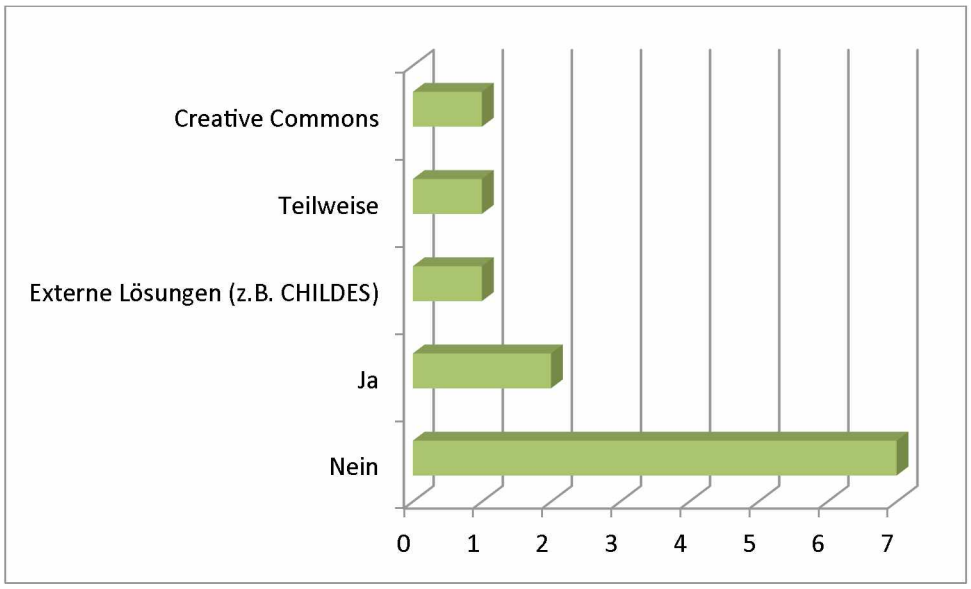
Ein Indikator für die Komplexität einer umfassenden LZA-Strategie ist die Garantie der Replizierbarkeit von Forschungsdaten, der durch eine Archivierung der Versionsgeschichte des Datenmanagements entsprochen werden kann. Sieben der befragten Einrichtungen haben bisher noch keine solche Lösung umgesetzt. Bei vieren ist dies bereits der Fall, ein weiteres befindet sich momentan in der Aufbauphase.

Frage 16: *Verfolgen Sie im Rahmen der Replizierbarkeit von Forschungsergebnissen Zugriffsstrategien?*



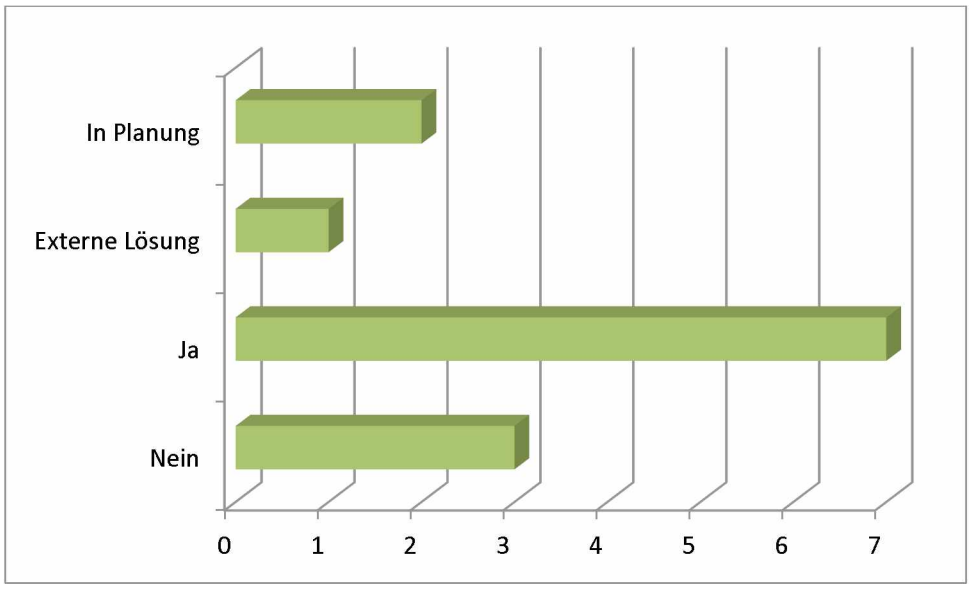
Analog zum Stand bei der Umsetzung von Archivlösungen hat die überwiegende Zahl der Befragten noch keine Zugriffsstrategien zum Zwecke der Replizierbarkeit ihrer Forschungsdaten – zu denken wäre etwa an ein Versionierungssystem – umgesetzt. Drei der Befragten haben individuelle Ansätze über Virtuelle Forschungsumgebungen (digitale Arbeitsplattform zur offenen Integration von Tools und Diensten sowie zur kollektiven netzbasierten Bearbeitung von Forschungsdaten), DOI bzw. ein Versionierungssystem.

Frage 17: *Gibt es Richtlinien zur Zitierung von Datensätzen in Publikationen?*



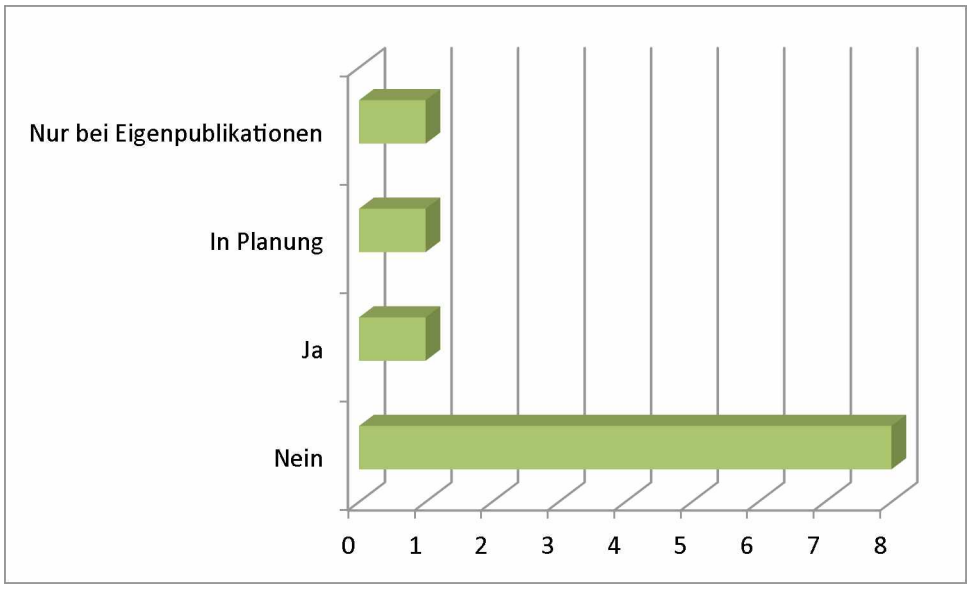
Bisher werden die Erhebung, Pflege und Archivierung von Forschungsdaten nur selten als vollwertige wissenschaftliche Leistung anerkannt, obgleich sich der Aufwand bis zu ihrer Anwendung und – zum Zwecke der Replizierbarkeit – darüber hinaus zu ihrer Aufbewahrung bisweilen sehr komplex gestaltet und kontinuierlich eigenständige, begleitende Forschungsarbeit erfordert. Um dies gerade auch vor dem Hintergrund zunehmender Evaluierung der Arbeit von Wissenschaftlern angemessen zu würdigen, empfiehlt sich eine einheitliche Systematik zur Zitation von Forschungsdaten von Nutzerseite. Überdies kann auf diese Weise die empirische Basis von wissenschaftlichen Auswertungen besser nachverfolgt werden. In diesem Punkt stellt sich die Situation in der erfassten Forschungslandschaft recht heterogen dar: sieben Befragte haben keine eigenen Lösungen implementiert, bei zweien ist es der Fall, die übrigen koppeln sich an externe Standards an, haben existierende Strategien nur bedarfsweise umgesetzt oder stellen alle Daten grundsätzlich unter den Maßgaben on Creative Commons frei zur Verfügung.

Frage 18: *Gibt es einen institutsweiten Dokumentenserver?*



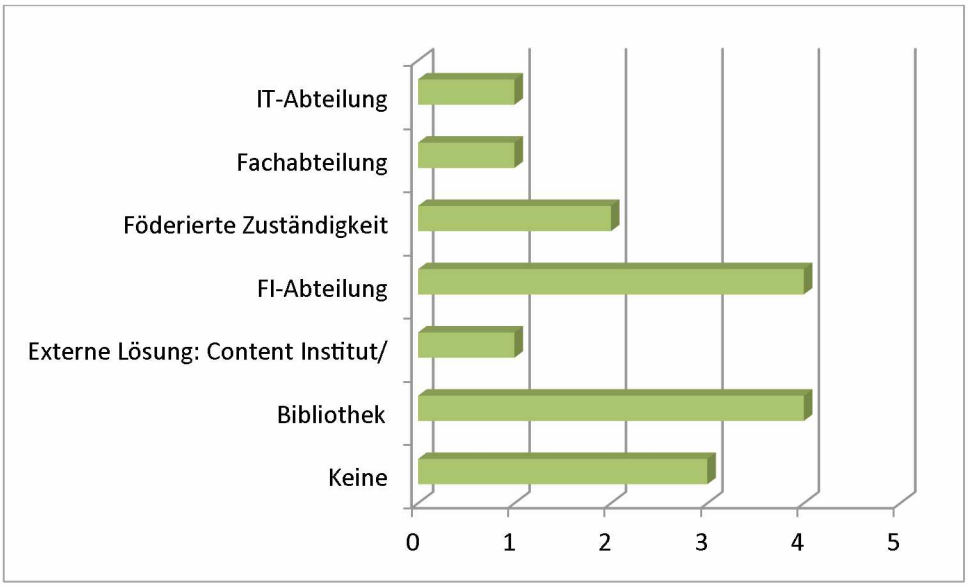
Zur einheitlichen und zentralen Dokumentation wissenschaftlicher Leistungen einer Forschungseinrichtung kann ggf. ein institutsweiter Dokumentenserver dienen. Im Rahmen von Evaluierungen, im Sinne einer nachhaltigen Archivierung oder bei der Erstellung von Berichten entstehen hierbei entscheidende Vorteile. Demzufolge haben auch sieben Einrichtungen ein solches Hilfsmittel aufgesetzt, zwei beabsichtigen dies für die Zukunft, wovon eines wiederum auf externe Lösungen zurückzugreifen plant.

Frage 19: *Gibt es eine Abgabepflicht für elektronische Belegexemplare von Publikationen der Mitarbeiter?*



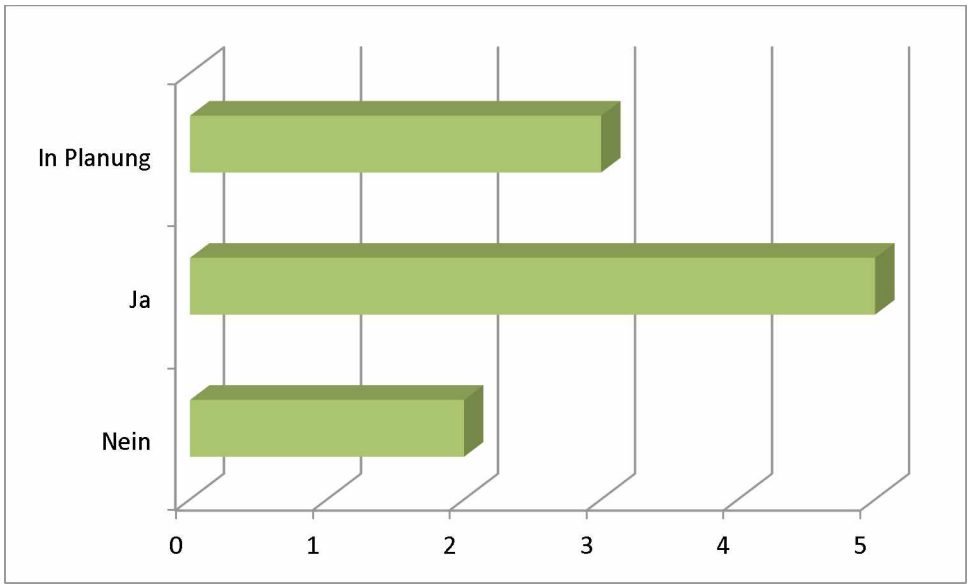
Um zentrale Verzeichnisse von Publikationen in elektronischer Form zuverlässig verwalten zu können, erscheint die automatische Abgabe eines Belegexemplars jedweder wissenschaftlichen Publikation eines Mitarbeiters zum Zwecke der Dokumentation sinnvoll. Die überwiegende Mehrheit der Befragten hat diesen Pfad bisher nicht beschritten. Ein Institut hat bereits eine vergleichbare Systematik umgesetzt, eines erwägt dies, eines beschränkt dies lediglich auf Publikation im Hausverlag.

Frage 20: *Welche Organisationseinheit ist für den Dokumentenserver zuständig?*



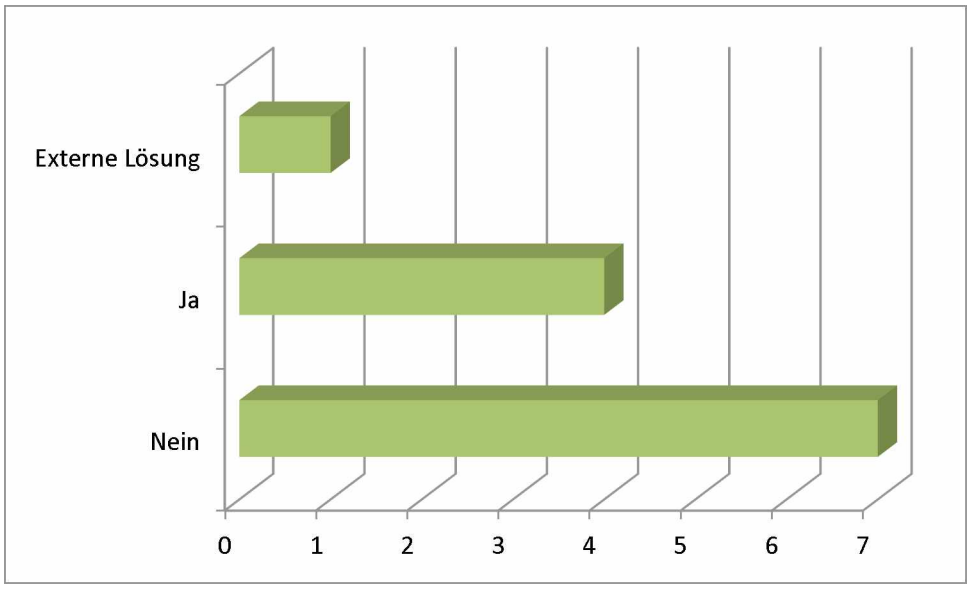
Für die Zuständigkeiten für einen zentralen Dokumentenserver lässt sich bei den Befragten kein einheitlicher Lösungsansatz erkennen: vier der Befragten siedeln diese Aufgabe bei der Bibliothek an, vier beteiligen ihre übergreifenden Abteilung für Forschungsinfrastrukturen, wobei sich die Zuständigkeit zwischen dieser und einer weiteren Abteilung aufteilt (IT- bzw. Fachabteilung). Erneut wählt ein Institut aufgrund personeller Engpässe externe Lösungen.

Frage 21: *Gibt es eine Open Access Strategie des Instituts?*



Innerhalb der Forschung ist es inzwischen weitgehend akzeptiert, gerade mit öffentlichen Geldern finanzierte Erkenntnisse oder Daten zumindest zu nicht kommerziellen Zwecken frei zur Verfügung zu stellen. Zu diesem Zweck existieren verschiedene Abstufungen des Open Access. Von den Befragten verfolgt die überwiegende Mehrheit eine solche Strategie oder entwickelt sie derzeit. Zwei Einrichtungen sehen aufgrund mangelnder Kapazitäten oder rechtlicher Bedenken derzeit noch keinen weg für eine vollständige Erfassung.

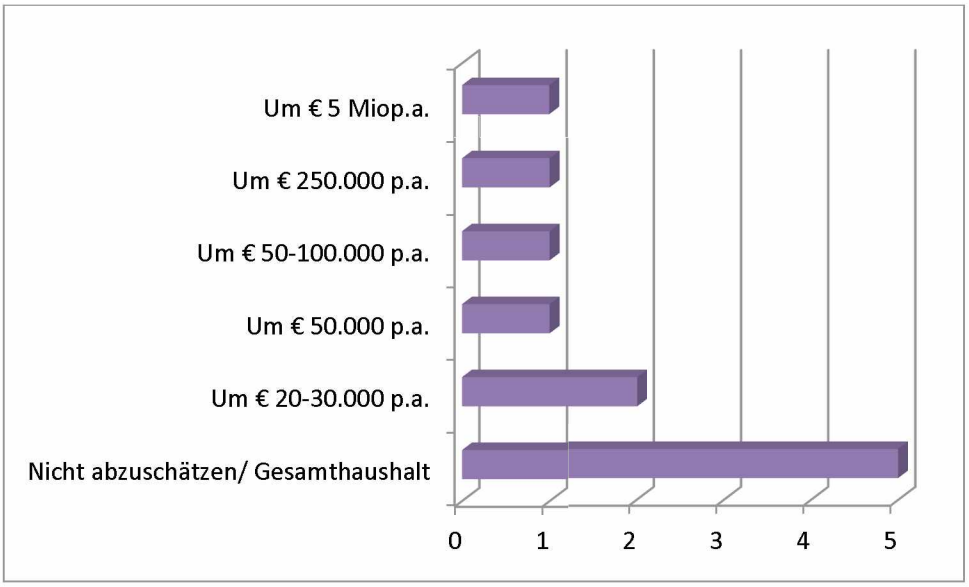
Frage 22: *Pflegen Sie ein Verzeichnis von Publikationen, die sich auf Forschungsergebnisse Ihres Instituts beziehen?*



Ergänzend zu Maßgaben der Zitation von Forschungsdaten empfiehlt sich die Dokumentation des Nutzungsgrades institutseigener Daten, um gerade im Rahmen von Evaluierungen einen maßgeblichen Impact der eigenen Ergebnisse in der Wissenschaft vorzeigen zu können. Die meisten Einrichtungen verzichten momentan noch auf solche Ansätze; einer der vier Befragten, die die Anwendung ihrer Daten nachverfolgen, sucht hierfür externe Lösungen.

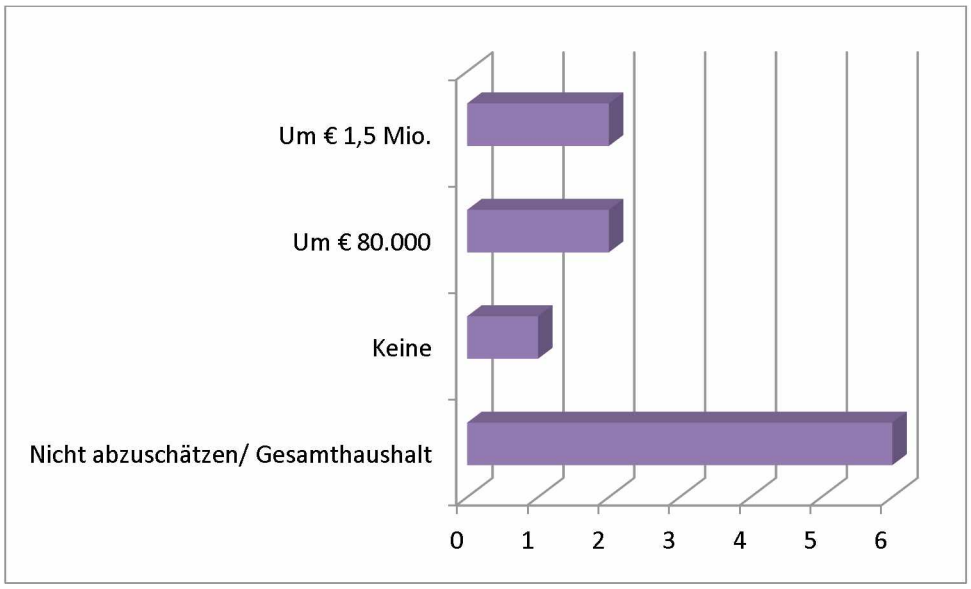
7.4 Kosten

Frage 23: *In welchem Rahmen bewegen sich laufende Kosten für das Forschungsdatenmanagement?*



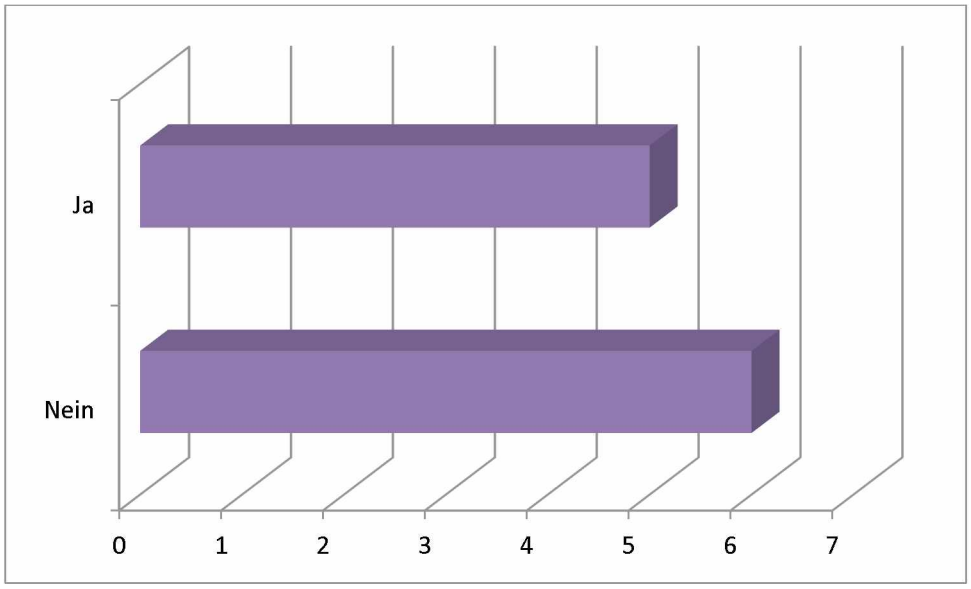
Für die künftige Planung von institutionellen Haushalten und die weitere Förderung von datenverarbeitenden Forschungsprojekten seitens der öffentlichen Hand wird die Kalkulation der entstehenden Kosten für den Betrieb eines umfassenden Forschungsdatenmanagements immer wichtiger. Nichtsdestotrotz stellt sich die genaue Abschätzung dieser Aufwände immer noch als recht schwierig dar, sodass bisher nur wenige orchestrierte Versuche in Verbundprojekten unternommen wurden (TextGrid, DARIAH, C3 usw.). So können fünf der Interviewten ihre Ausgaben kaum abschätzen, da sie aus dem Gesamthaushalt bestritten werden. Bei den übrigen Instituten bewegen sich die Kosten je nach Größe und Umfang des Datenbestandes zwischen € 20.000 und € 5 Mio. im Jahr.

Frage 24: In *welchem Rahmen bewegen sich investitive Kosten?*



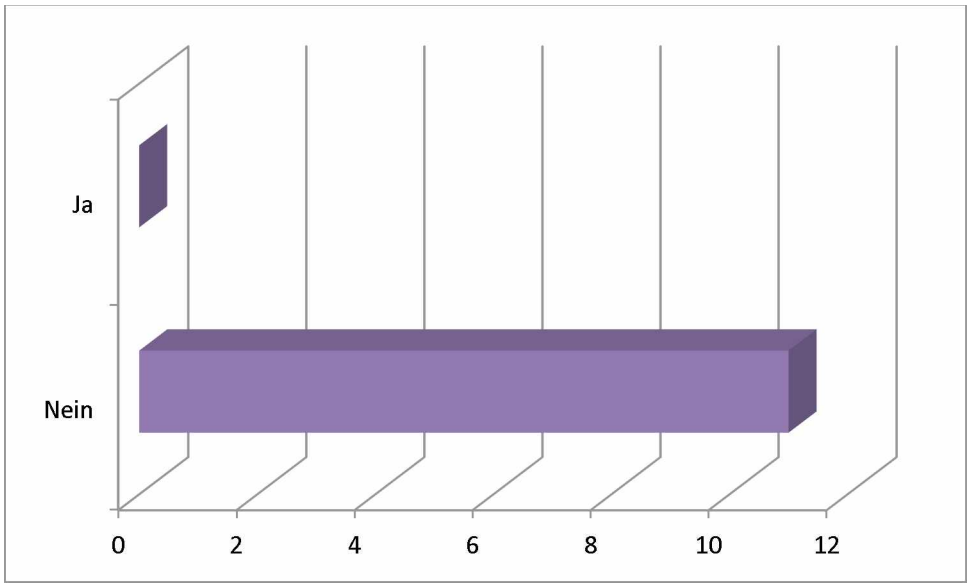
Analog hierzu verhält sich die Situation bei der Kalkulation der Beschaffungskosten für Hard- oder Software und zugehöriger Infrastruktur. Sechs Befragte konnten diese Fragen nicht beantworten; ein Institut hat in diesem Punkt keine Aufwände, weil externe Lösungen kostenneutral genutzt werden. Die übrigen investieren bis € 1,5 Mio. in ihre Infrastruktur.

Frage 25: *Gibt es eine eigene Kostenstelle für Personalkosten zum Betrieb von IT-Systemen (z.B. für Systemadministratoren)?*



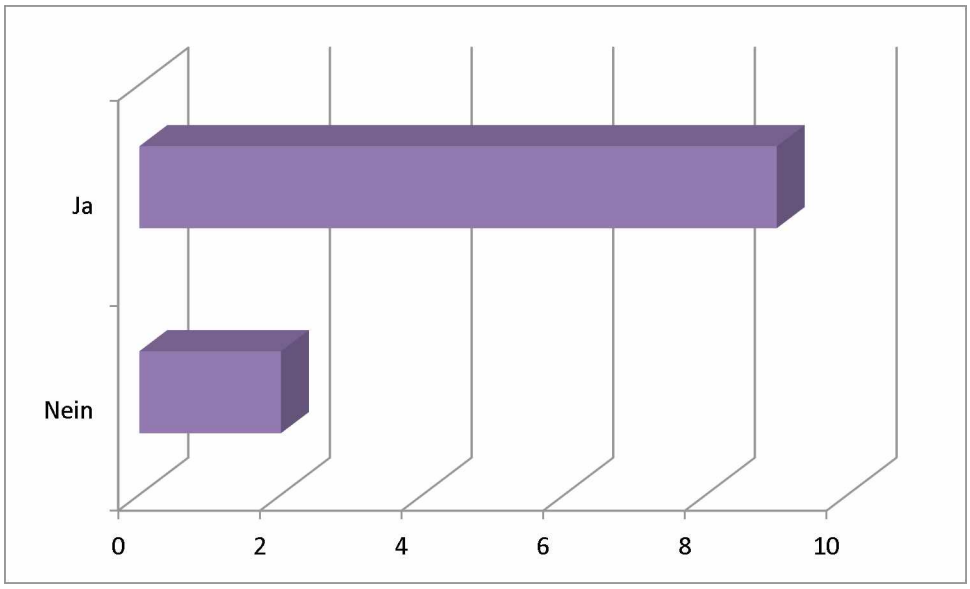
Begründet sind die Schwierigkeiten bei der Kostenschätzung meist in der Tatsache, dass für Einzelposten im Umfeld der IT-Beschaffung, des Betriebs und des Datenmanagements nicht gesondert erfasst werden: Sechs Einrichtungen haben keine Kostenstelle für Personalmittel ausgewiesen, während die übrigen dies zumeist im Rahmen ihrer IT-Abteilungen tun.

Frage 26: *Gibt es eine eigene Kostenstelle für Energiekosten zum Betrieb der IT-Infrastruktur (z.B. den Strom, den die IT-Komponenten und deren Klimatisierung verbrauchen) oder geht dies in die Energiekosten des gesamten Instituts ein?*



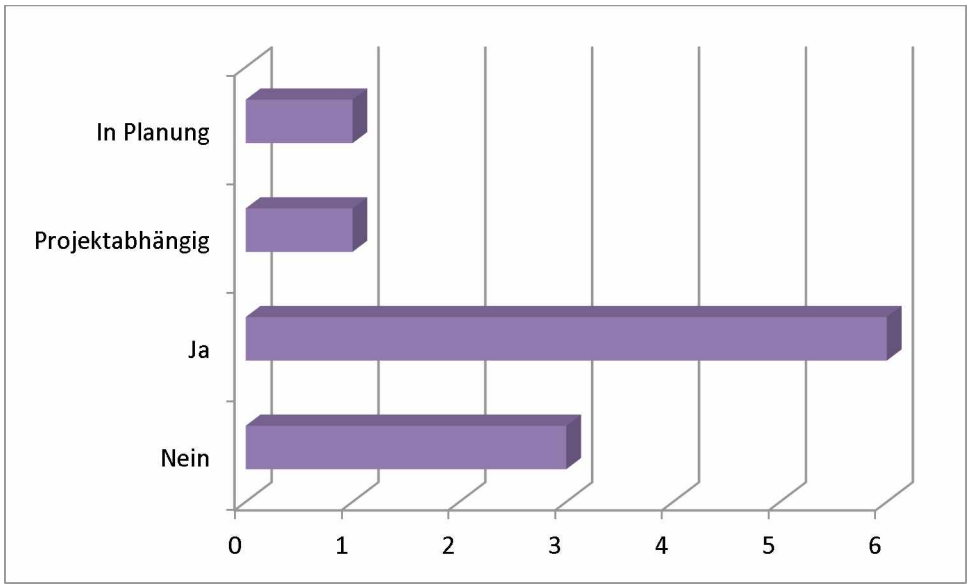
In zunehmendem Maße werden – vornehmlich für Hochschulen, aber tendenziell auch im Rahmen der Förderung von kostenintensiven Forschungsprojekten – Überlegungen zur verpflichtenden Angabe von Vollkostenrechnungen bei der Beantragung von Fördergeldern laut. Einen nicht unerheblichen Anteil bilden hierunter nicht zuletzt hinsichtlich jüngster Verwerfungen auf dem Strommarkt die Energiekosten für den Betrieb der IT-Infrastruktur. Keines der befragten Institute hat jedoch einen gesonderten Posten für Energiekosten ausgewiesen.

Frage 27: *Gibt es Rahmenverträge für die Beschaffung oder den Support mit Soft- oder Hardwareanbietern?*



Für die gesammelte Bestellung von Soft- oder Hardware bzw. deren Support lassen sich Kosten einsparen, wenn keine Insellösungen verfolgt, sondern über einen Rahmenvertrag standardisierte Pakete eingekauft werden. So verfolgen auch die meisten Interviewpartner diesen Weg. Zwei der Befragten geben Ausschreibungen heraus oder sehen hinsichtlich ihres Bedarfs keine Notwendigkeit für Rahmenverträge.

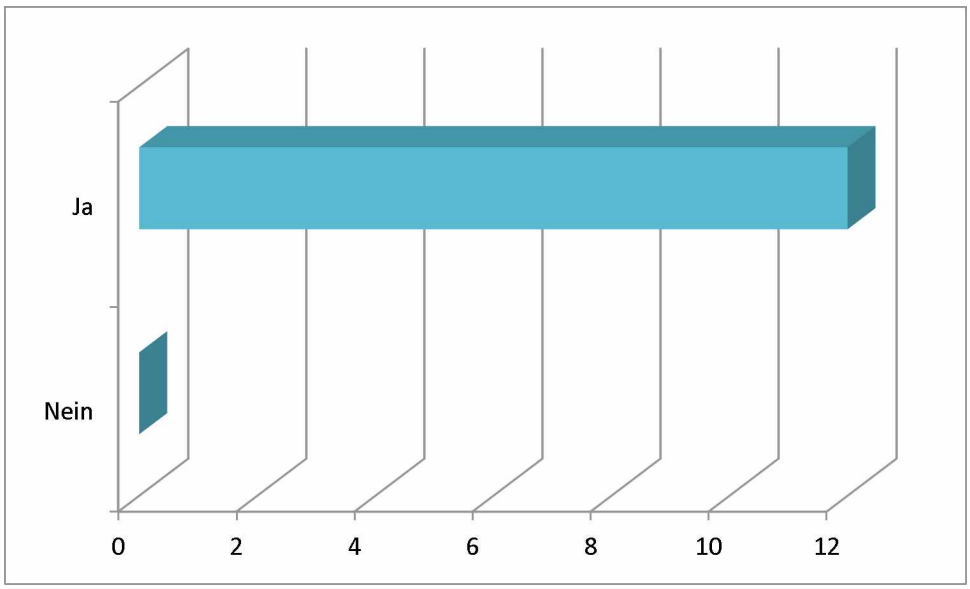
Frage 28: *Gibt es inter-institutionelle Kooperationen zur gemeinsamen Hardwarenutzung?*



Zusätzlich zu Rahmenverträgen schließen sich viele Einrichtungen zu Clustern bei der Beschaffung von Hard- und Software bzw. Support zusammen oder beteiligen sich an Beschaffungen größerer bzw. übergeordneter Institutionen. Sieben Befragte beantworteten diese Frage positiv, wovon sich eines derzeit noch in der Planungsphase befindet, während drei eigenständige Lösungen suchen. Ein Institut wählt übergreifende Ansätze nicht hausintern, wohl aber für Neuanschaffung auf Projektebene.

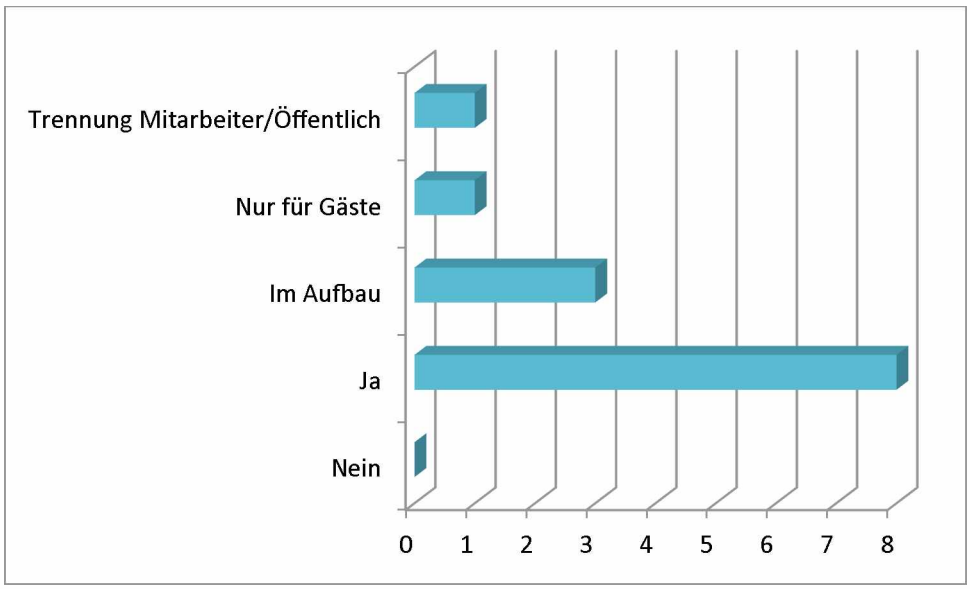
7.5 Dienste

Frage 29: *Stehen Basisdienste wie eMail, WWW u.Ä. zur Verfügung?*



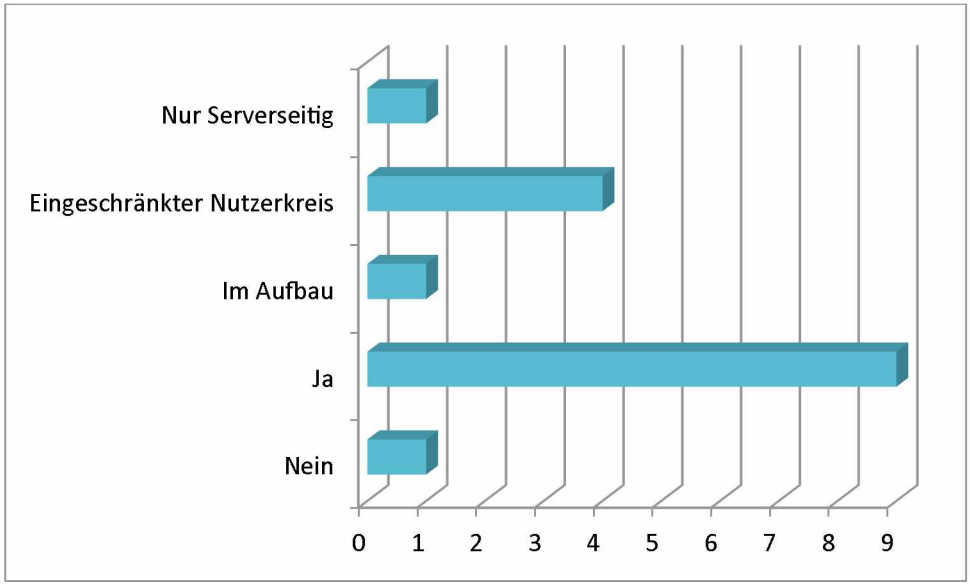
Die Arbeit und Kommunikation mit dem und im Internet stellt inzwischen eine solche Selbstverständlichkeit dar, dass sich kein Institut diesem Ansatz entzogen hat.

Frage 30: *Steht WLAN zur Verfügung?*



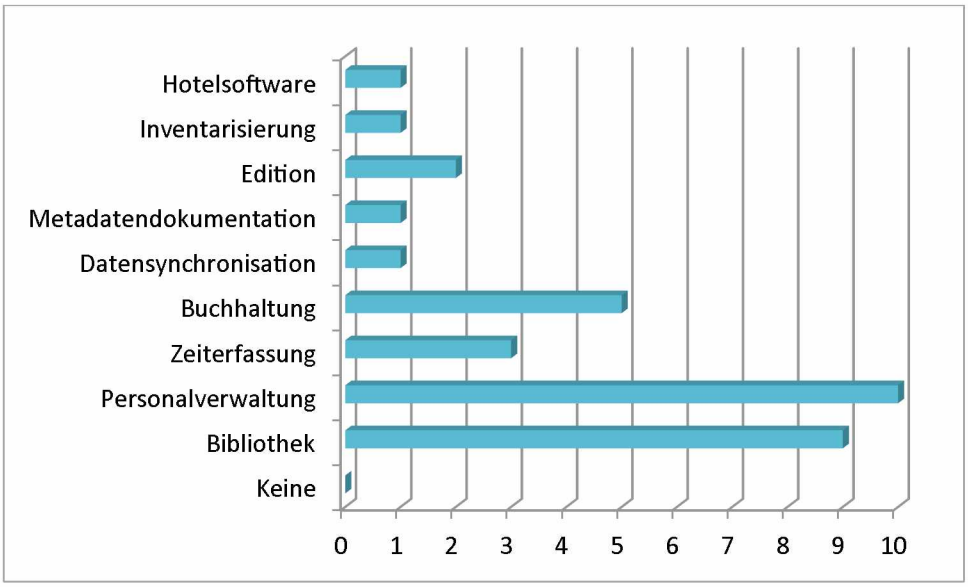
Fast so eindeutig stellt sich die Situation mit Blick auf die Nutzung von WLAN in den einzelnen Häusern dar. Zwar bietet die überwiegende Zahl der Befragten ihren Mitarbeitern und Gästen diesen Dienst an (acht von zwölf) oder baut ihn derzeit auf (drei). Doch beschränken zwei Institute den Zugang für Nichtmitarbeiter bzw. behalten ihn ausschließlich Gästen vor.

Frage 31: *Steht VPN zur Verfügung?*



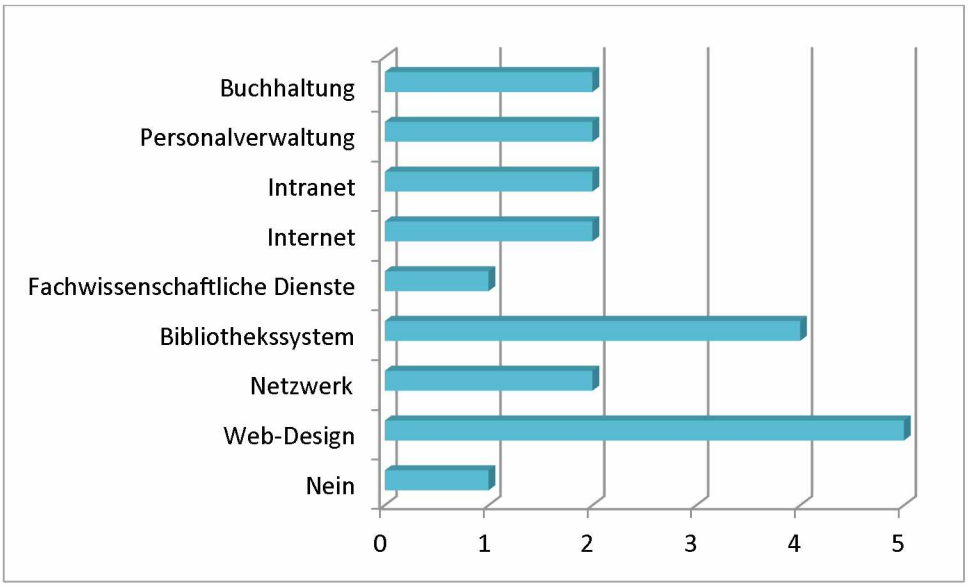
Die steigende Zahl von in Verbünden durchgeführten Vorhaben erfordert eine zunehmende Mobilität der Wissenschaftler. Um den Mitarbeitern dennoch Zugriff auf institutsspezifische Dienste und Ressourcen (eMail, Intranet usw.) zu ermöglichen, sind externe Zugangsformen zu Institutsnetzwerk unerlässlich. Will man sich nicht auf externe Angebote verlassen, stellt ein hauseigenes VPN eine optimale Alternative dar. Zehn der zwölf Interviewpartner haben VPN in ihren Häusern inzwischen eingerichtet bzw. bauen es gerade auf – vier davon schränken dieses Angebot auf einen bestimmten Nutzerkreis ein (zumeist Leitungsebene). Nur zwei Institute bieten kein generelles bzw. nur serverseitiges VPN an.

Frage 32: *Welche Spezialdienste stehen zur Verfügung?*



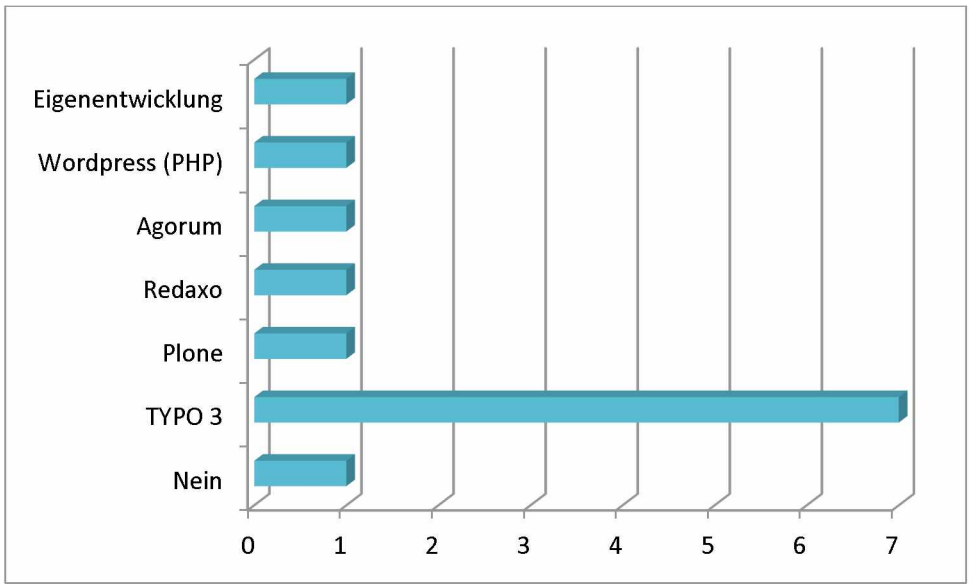
Neben den bisher angesprochenen Diensten betreiben viele Institute gesondert auf ihre organisatorischen Bedarfe bzw. disziplinären Fragestellungen angepasste fachliche oder administrative Spezialdienste. Durchweg vorhanden sind ein Personalverwaltungs- und – bis auf eine Ausnahme – ein Bibliothekssystem. Viele der Befragten (fünf) wickeln auch ihre Buchhaltung bzw. Kosten-Leistung-Rechnung (drei), die gerade hinsichtlich der Evaluierung von Forschungsinstitutionen zunehmend an Bedeutung gewinnen, über gesonderte Dienste ab. Darüber hinaus existieren spezialisierte Systeme zur Edition von Forschungsergebnissen (zwei), Datensynchronisation, Metadatendokumentation und Inventarisierung. Eine Infrastruktureinrichtung verfügt sogar über umfassende Unterbringungsmöglichkeiten für Gäste, die über ein Hotelverwaltungssystem koordiniert werden.

Frage 33: *Erfolgt eine Ausgliederung einzelner Dienste oder Dienstleistungen?*



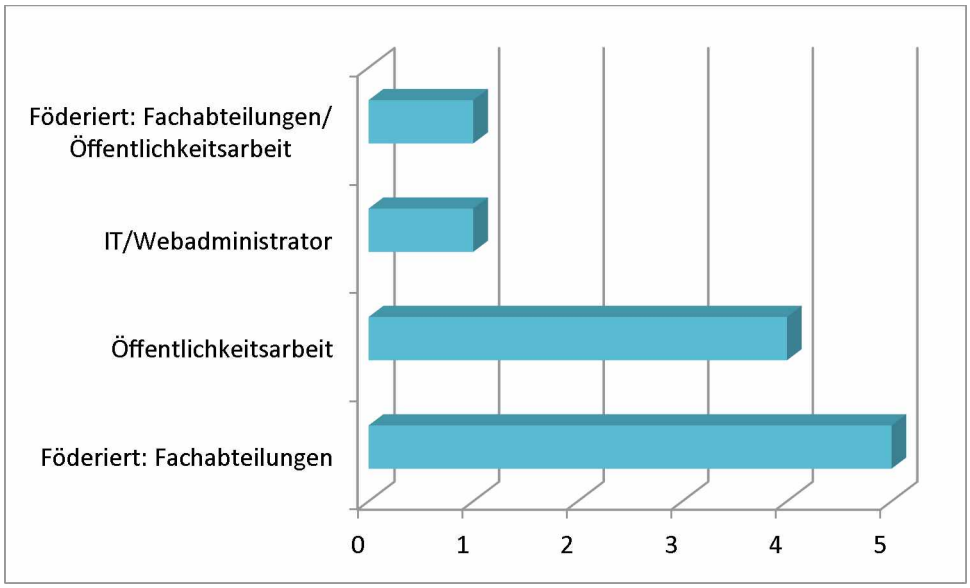
Wie bereits zu beobachten war, suchen viele Institutionen zumeist aus Kapazitätsgründen externe Dienstleister – sei es ein kommerzieller Anbieter, sei es eine akademische Partnerorganisation oder übergeordnete Einrichtung – zum Betrieb und Support bestimmter Infrastrukturkomponenten. Manchmal wird auf Poollösungen zurückgegriffen. Die Dienste, die am häufigsten nicht von den Instituten selbst betreut, sondern zumeist an kommerzielle Anbieter ausgelagert werden, sind das Design und gelegentlich auch die Pflege der Institutswebseite (fünf von zwölf). Weitaus weniger häufig werden Poollösungen für das Bibliothekssystem genutzt (drei). Andere Auslagerungen betreffen Hosting und Support des Internetdienstes (zwei), des Intranets (zwei), des hauseigenen Netzwerks (zwei), der Buchhaltung (zwei), Personalverwaltung (zwei) sowie fachlicher Spezialdienste (zwei) u.A.

Frage 34: *Verwenden Sie ein Content Management System und, wenn ja, ggf. welches?*



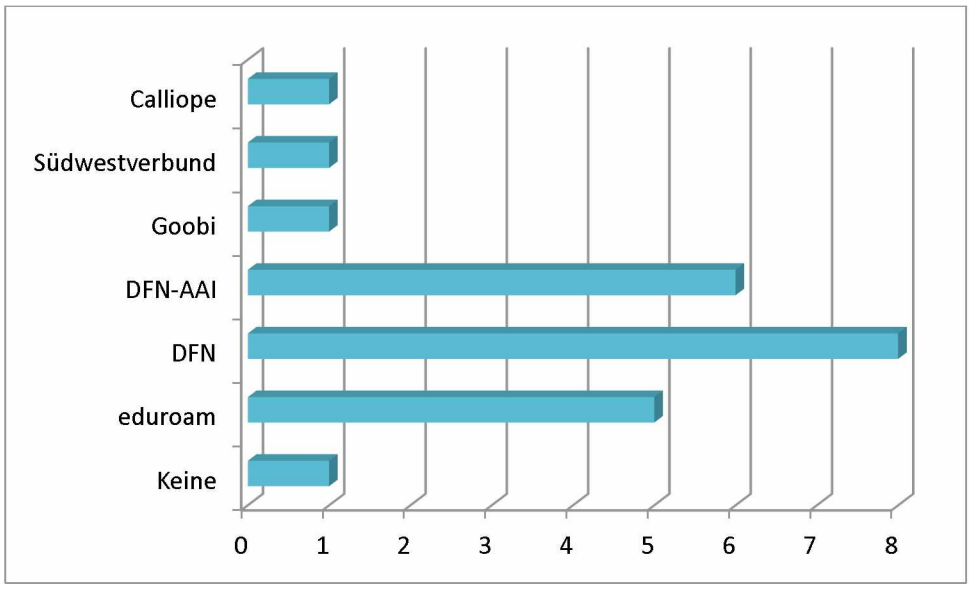
Trotz der relativ häufigen Ausgliederung von Hosting oder Design ihres Internetauftritts pflegen bis auf ein Institut, das die Pflege seiner Homepage vollständig extern vergibt, alle Befragten die so nach außen getragenen Inhalte über ein Content Management System selbst. Lediglich eines der interviewten Häuser hat zu diesem Zweck ein eigenes System entwickelt. Alle anderen bedienen sich Standardlösungen, unter denen TYPO3 das mit Abstand häufigste ist (sieben von zwölf).

Frage 35: *Wer pflegt die Inhalte im Content Management System?*



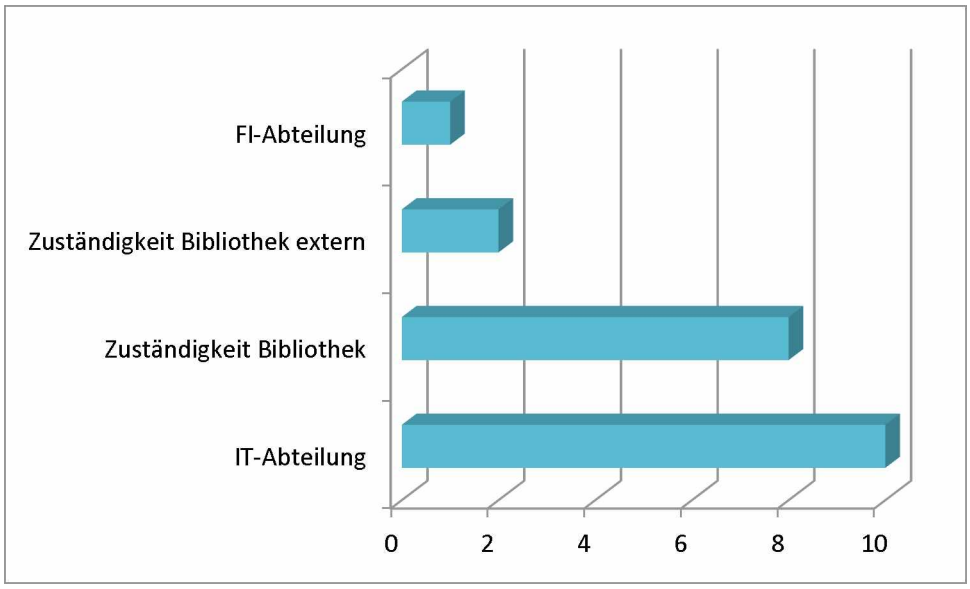
Bei den Institutionen, die ihre Inhalte selbst pflegen, ist die Zuständigkeit für die Inhalte verteilt. Zumeist sind die einzelnen Fachabteilungen selbst für die Fütterung mit Informationen verantwortlich – ein Institut verfügt zwar nicht über ein CMS (vgl. Frage 34), pflegt seine Homepage dennoch selbst. Sehr häufig obliegt diese Pflicht auch der Öffentlichkeitarbeit (vier); in einem Falle ist die Pflege derart dezentral organisiert, dass unterschiedliche Aufgaben auf die Fachabteilungen, die ausschließlich fachliche Informationen einpflegen, die Öffentlichkeitsarbeit, welche allgemeine Information und das Design beisteuert, sowie die IT-Abteilung mit ihrer technischen Expertise verteilt werden.

Frage 36: *An welchen Verbünden nehmen Sie teil/wo sind Sie Mitglied?*



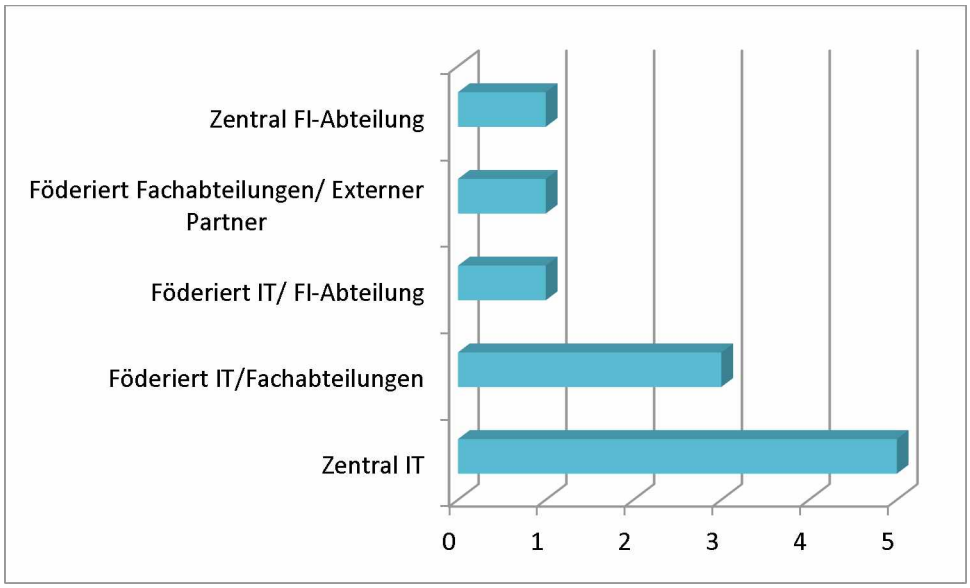
Innerhalb der Wissenschaftslandschaft haben sich inzwischen einige Verbünde aus akademischen Betreibern etabliert, die Partnerinstitutionen mit einem breiten Spektrum an Online-Diensten versorgen können. Dazu zählen insbesondere Roaming, Accountverwaltung oder Telefoniedienste. Recht weit verbreitet ist die (assoziierte) Mitgliedschaft im DFN e.V. (acht von zwölf). Einige dieser Mitglieder nutzen auch dessen Authentifizierungs- und Authorisierungsinfrastruktur, um Zugriff auf geschützte Informationen auf verteilten Webservern zu ermöglichen. Fünf Einrichtungen Nutzen überdies auch die von *eduroam* angebotenen Dienste, die allen Teilnehmern einen Webzugang an den beteiligten Standorten bieten. Seltener sind die Teilnahme an dem Digitalisierungsverbund *Goobi* für Bibliotheken oder der vollständige Verzicht auf eine Verbundmitgliedschaft.

Frage 37: *Wer ist für den End-User-Support der Arbeitsplätze zuständig (evtl. auch Bibliothek)?*



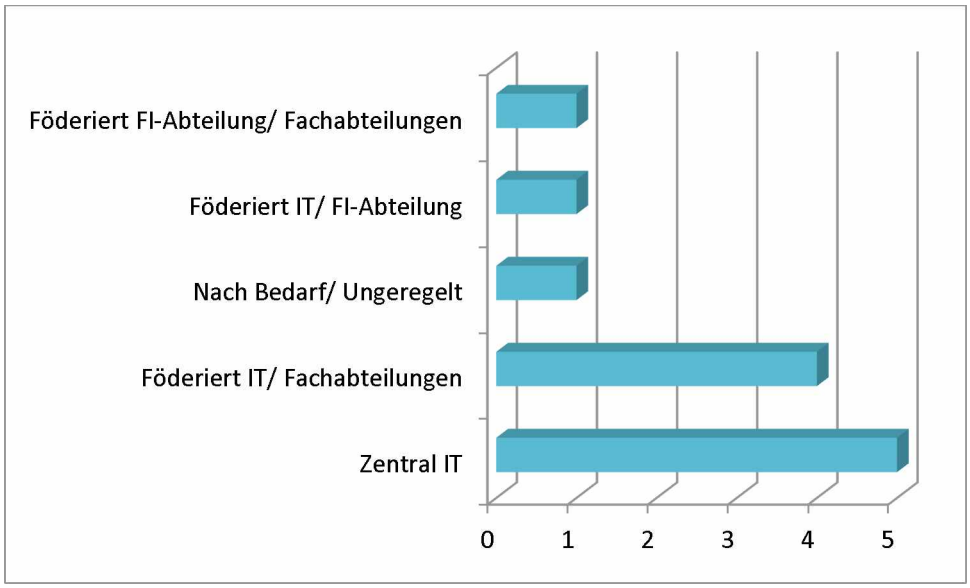
Den End-User-Support organisiert überwiegend die hauseigene IT-Abteilung. Ein Institut hat zwar die meisten Belange der EDV-Betreuung ausgelagert, benennt jedoch aus dem wissenschaftlichen Stab einen IT-Beauftragten, der die Arbeitsplätze vor Ort betreut. Gleichzeitig wird jedoch in diesem und in einem weiteren Institut der Support für Bibliotheksrechner extern vergeben. Viele der Befragten (acht) trennen die Zuständigkeiten zwischen Arbeitsplatz- und Bibliotheksrechnern. Die IT-Abteilung betreut mit einer Ausnahme auch in den befragten Infrastruktureinrichtungen hierbei die Rechner der wissenschaftlichen Mitarbeiter, während der Support der oft öffentlichen PCs der Ausleihen vor Ort erfolgt.

Frage 38: *Wer ist für den Systembetrieb zuständig?*



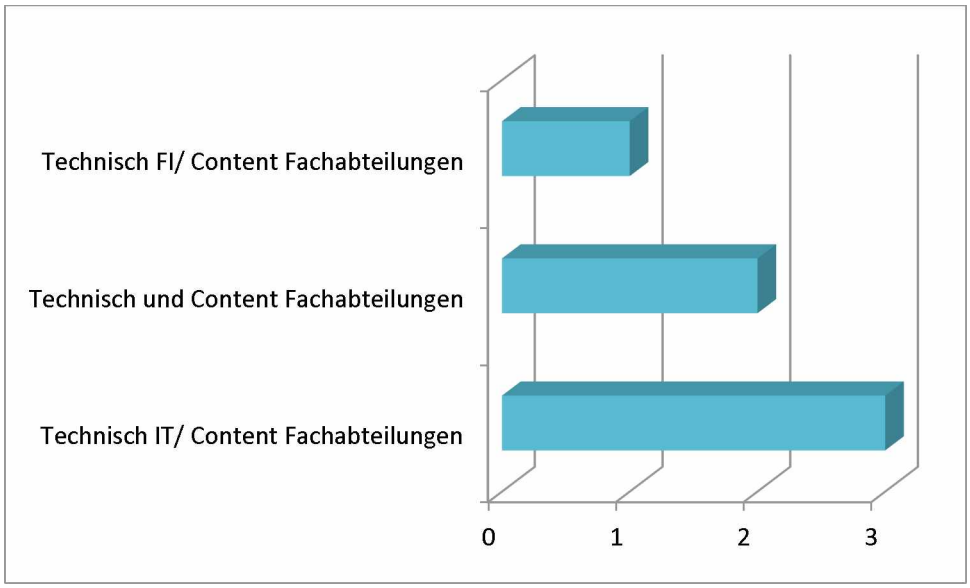
Der Betrieb von informationstechnischen Systemen aller Art (Software, Hardware, Netze) wird oft von der zentralen IT-Abteilung übernommen (fünf von zwölf). In machen Institutionen teilen sich diese Aufgabe sowohl die IT-Abteilung als auch einzelne Fachabteilungen, die mit den örtlichen PCs arbeiten. Nur in je zwei Fällen erfolgt diese Dezentralisierung zwischen der EDV einerseits und der zentralen Abteilungen für Forschungsinfrastrukturen bzw. einem externen Dienstleister andererseits; ein weiterer Einzelfall stellt die Übertragung dieser Verantwortlichkeit auf eine zentrale FI-Abteilung in einer Infrastruktureinrichtung dar.

Frage 39: *Wer ist für fachspezifische Dienste der Fachabteilungen zuständig?*



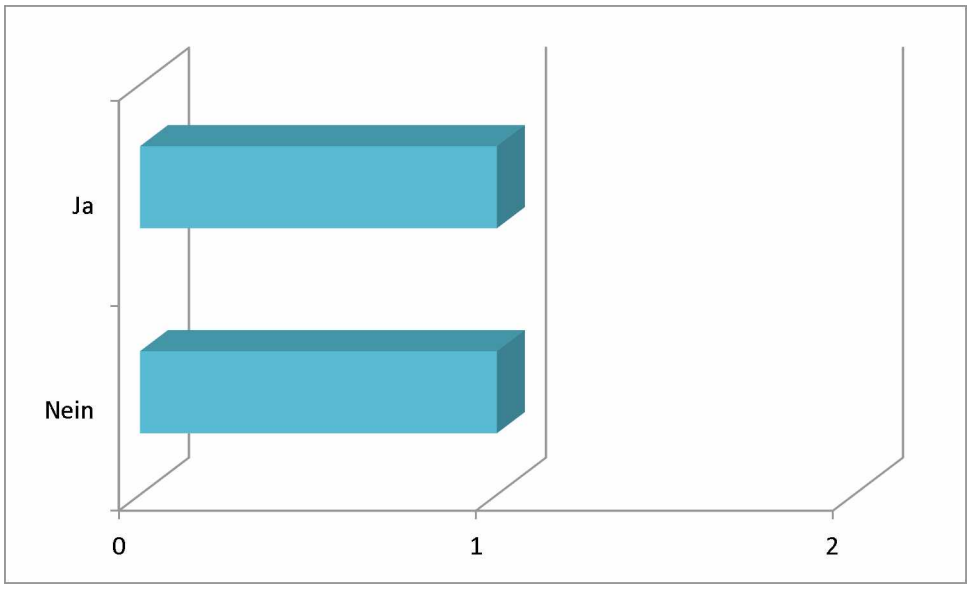
Oftmals werden in den Fachabteilungen spezialisierte Dienste betrieben, die auf das wissenschaftliche Arbeiten dieser Abteilung zugeschnitten sind. Auch in diesem Fall obliegt deren Betreuung oft der IT-Abteilung (fünf von zwölf). Nicht deckungsgleich mit der Zuständigkeit für den Systembetrieb (vgl. Frage 38) besteht in vier Häusern eine föderierte Zuständigkeit zwischen EDV und Fachabteilungen. Je ein Institut hat keine verbindliches Vorgehen für solche Fälle eingeführt, teilt diese Aufgabe zwischen IT und Forschungsinfrastrukturabteilung auf oder überantwortet sie einer zentralen IT-Abteilung.

Frage 40: *Wo liegen bei gemischter Zuständigkeit die Grenzen?*



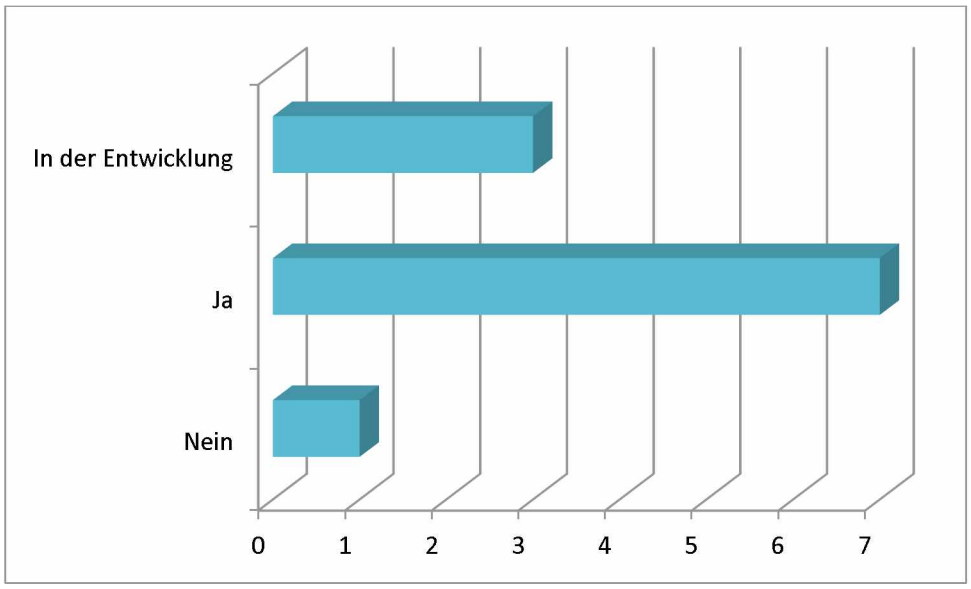
Bei föderierten Zuständigkeiten in vier Instituten und zwei Infrastruktureinrichtungen erfolgt eine Aufteilung der Zuständigkeiten zwischen rein technischer Betreuung fachspezifischer Dienste in der EDV/FI und der Pflege der Inhalte durch die Fachabteilungen selbst. In zwei Häusern erfolgt keine Differenzierung.

Frage 41: *Wenn die Dienste durch die Fachabteilungen betrieben werden: Gibt es Vorgaben, wie die Dienste betrieben werden sollen?*



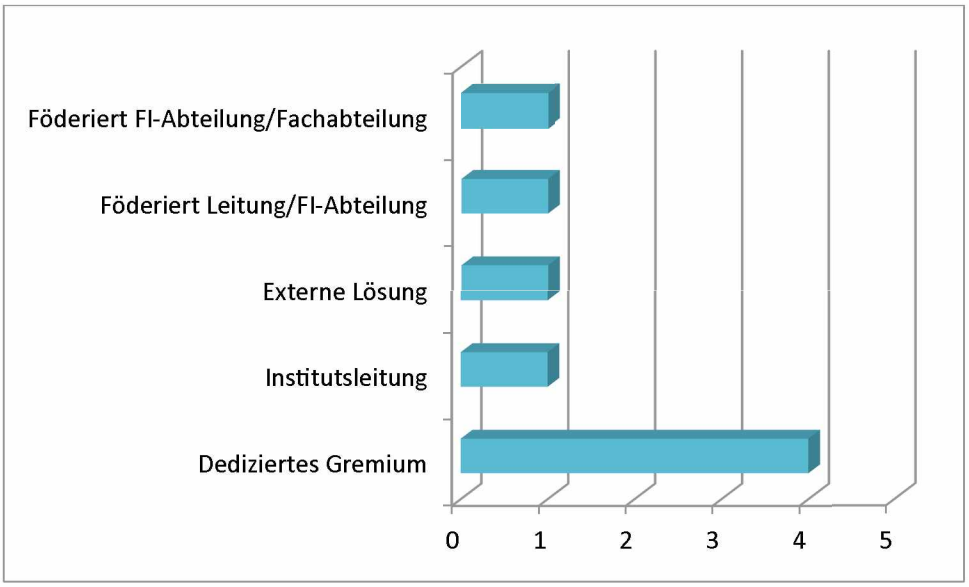
Von den beiden Instituten, die ihren Fachabteilungen viel Freiheit im Betrieb spezialisierter Dienste lassen, unterstützen die IT-Abteilungen ihre wissenschaftlichen Kollegen in einem Fall durch Vorgaben zur Art des Betriebs. Das andere Institut lässt seiner Abteilung für Forschungsinfrastruktur völlig freie Hand.

Frage 42: *Gibt es eine mittel- oder langfristige IT-Strategie in Ihrem Haus?*



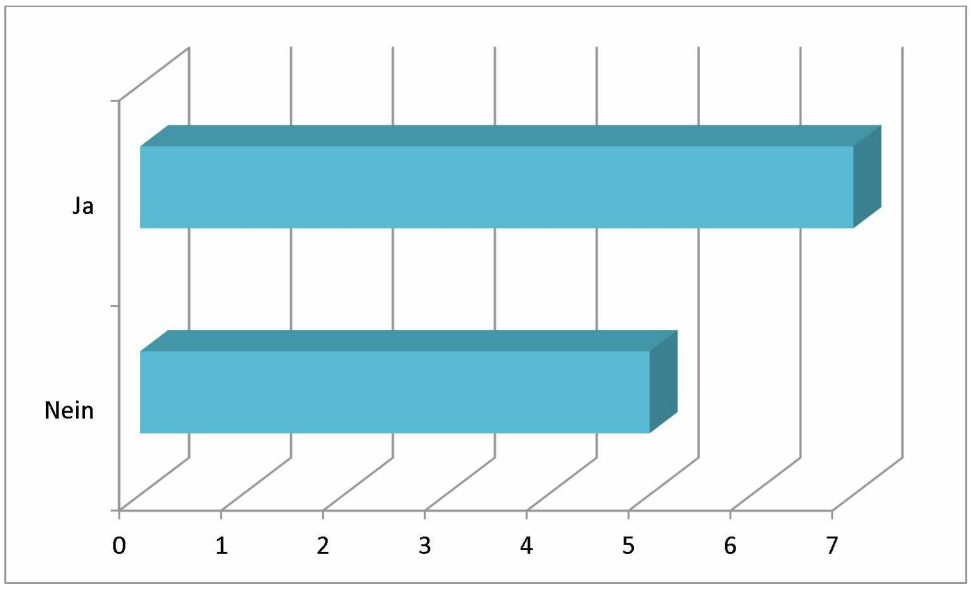
Die Frage nach der Umsetzung einer gesteuerten IT-Strategie, im Rahmen derer auf zentraler Ebene die Weichen für die Gesamtausrichtung des Instituts hinsichtlich Finanzierung von Ausgaben und Anschaffungen, Einsatz und Schwerpunkten der Arbeit mit IT, Standardisierung und Zusammenstellung von Soft- und Hardware, Angebot an Diensten, Sicherheits- und Datenschutzrichtlinien, Notfallplänen, Nachhaltigkeit und ggf. Auslagerung gestellt werden, beantworteten die Befragten zumeist positiv: Sieben von zwölf Einrichtungen haben einen solchen Plan erarbeitet und umgesetzt. Drei weitere entwickeln gerade entsprechende Maßgaben, während lediglich ein Institut aufgrund seiner geringen Größe noch nicht die Notwendigkeit für weitere Schritte sieht.

Frage 43: *Wer legt die IT-Strategie fest?*



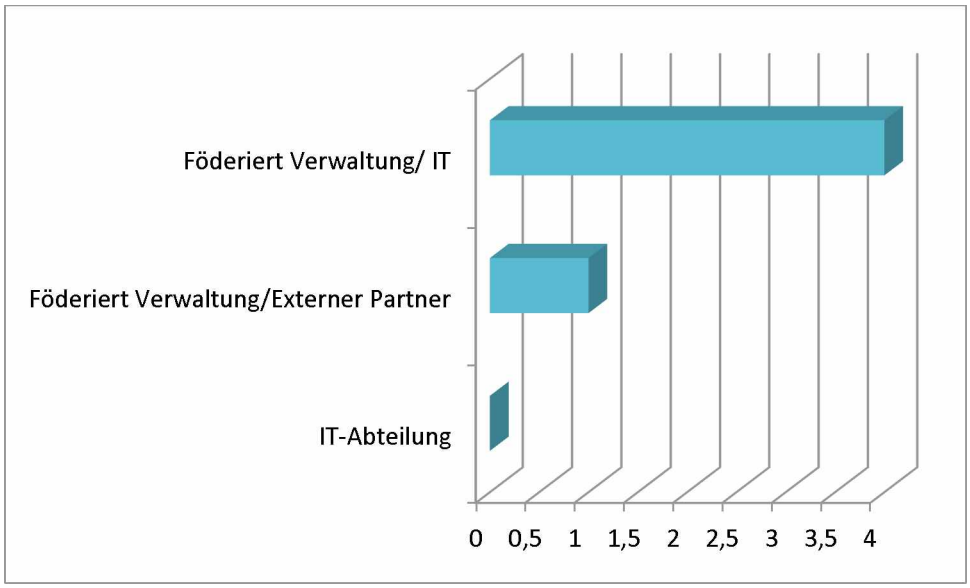
Für die Ausarbeitung der IT-Strategie haben viele Einrichtungen (vier von zehn mit [in Entwicklung befindlicher] IT-Strategie) ein speziell für diese Arbeit vorgesehenes Expertengremium eingesetzt, das sich zumeist aus Mitgliedern der IT-Abteilung, der Fachabteilungen, der Institutsleitung und der Verwaltung zusammensetzt. Je ein Institut wählt einen Top-Down-Ansatz, schließt sich externen Vorgaben an oder bevorzugt einen föderierten Entscheidungsweg zwischen bestehenden Fachabteilungen bzw. der FI-Abteilung und der Leitungsebene.

Frage 44: *Gibt es ein Identitätsmanagementsystem in Ihrem Haus?*



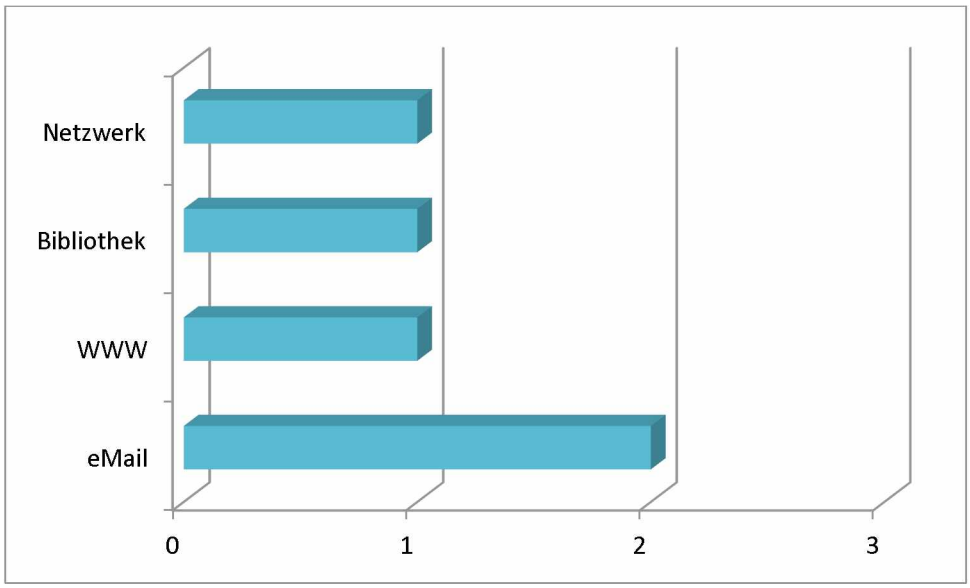
Zur Bündelung sämtlicher Zugangsdaten und -wege zu den in Instituten vorhandenen Ressourcen und Diensten von Forschungsdaten über Intranet bis hin zu eMail besteht die Möglichkeit der Einrichtung eines zentralen Identitätsmanagementsystems, mithilfe dessen über einen integrierten Account möglichst alle Angebote erschlossen und Zugriffsrechte gesteuert werden kann. Sieben der zwölf befragten Häuser verfügen über ein solches System, die übrigen vergeben individuelle Zugangswege.

Frage 45: *Wer verwaltet das Identitätsmanagementsystem?*



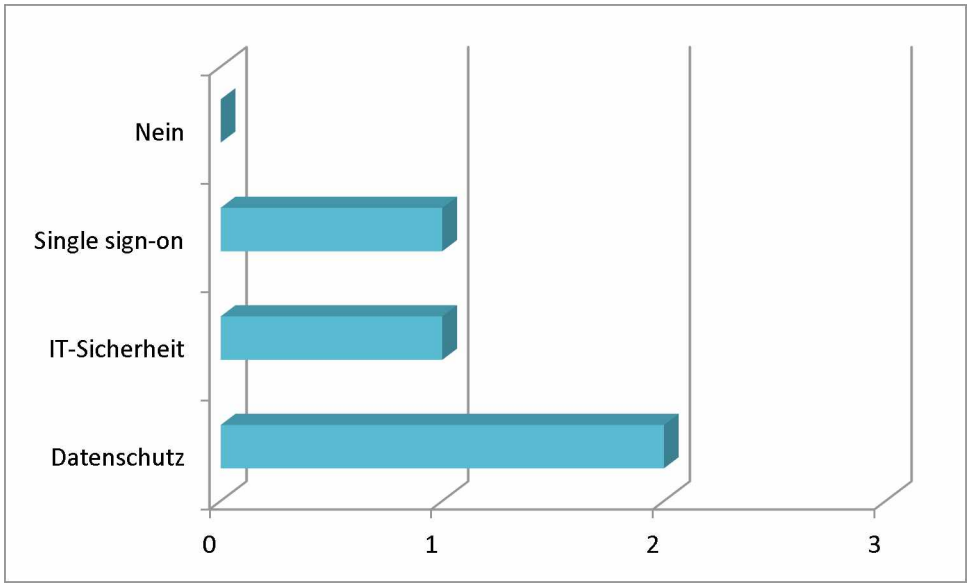
Um die zur Steuerung der Zugriffsrechte notwendigen Informationen sowie die technischen Voraussetzungen zusammenzuführen, kooperieren in fast allen Instituten (Personal-) Verwaltung und IT-Abteilung in der Pflege des Identitätsmanagementsystems. Ein Institut nutzt einen externen Service und stimmt sich über die Verwaltungsabteilung mit diesem Partner ab.

Frage 46: *Welche Dienste sind an das Identitätsmanagementsystem angeschlossen?*



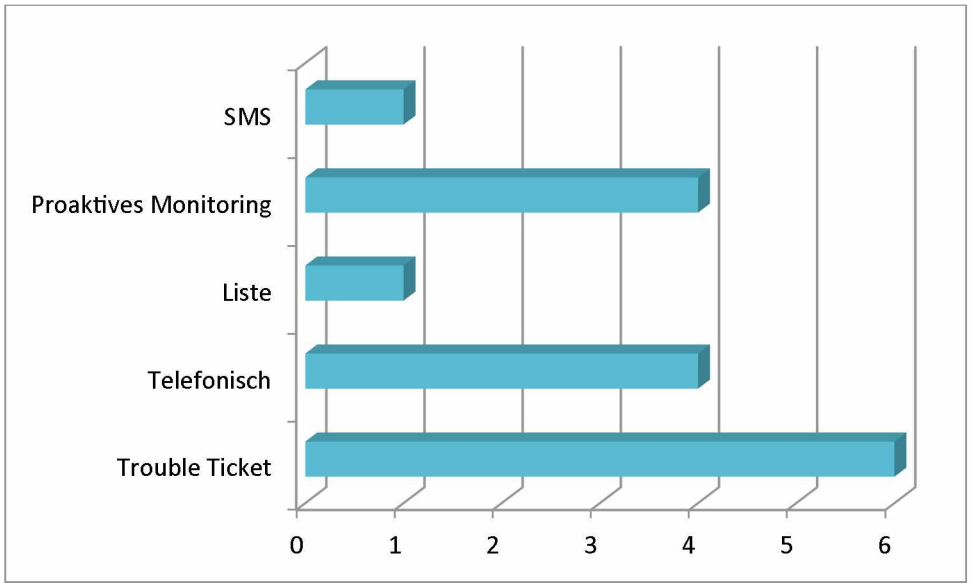
Zumeist ist es der Zugang zum E-Mail-Account der Mitarbeiter, der über das Identitätsmanagementsystem verwaltet wird. Weitere Dienste sind der Zugriff auf das Internet, die Nutzung von Bibliotheksdiensten und der allgemeine Zugriff auf das Institutsnetzwerk.

Frage 47: *Gibt es Kriterien für den Anschluss dieser Dienste an das Identitätsmanagementsystem?*



Ein wesentliches Kriterium dafür, den Zugriff über das Identitätsmanagementsystem zu steuern oder gar einzuschränken, stellt vor allem der Schutz insbesondere personenbezogener Daten dar. Aber auch die Sicherheit des Gesamtsystems bedingt Einschränkungen. Eine Infrastruktureinheit richtet hingegen alle im Identitätsmanagementsystem verwalteten Dienste so aus, dass sie durch einmalige Authentifizierung erreicht werden sollen (Single sign-on).

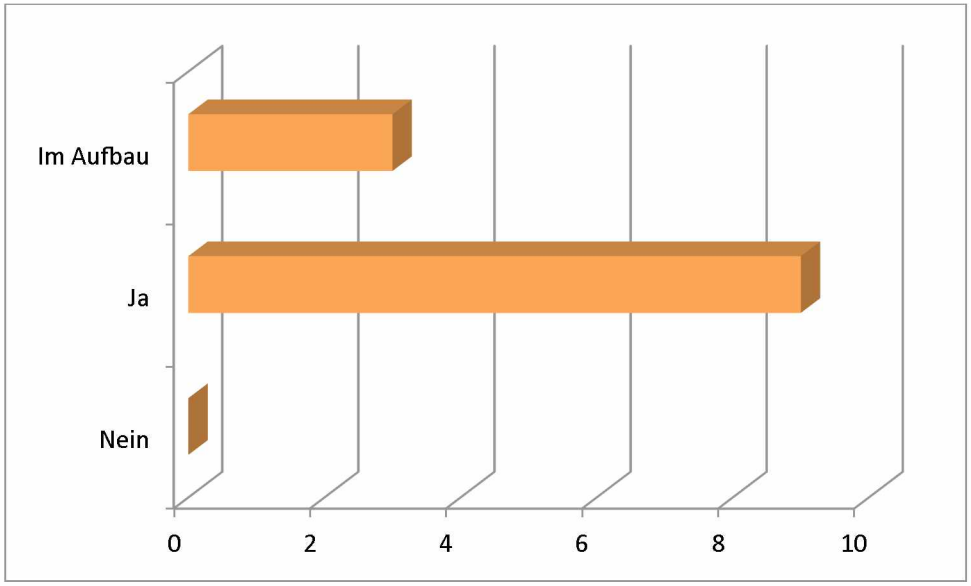
Frage 48: *Wie ist der IT-Support organisiert?*



Im Falle von Fehlfunktionen der Arbeitsplatzrechner, von netzbasierten Ressourcen oder Diensten sind einzelne Wissenschaftler zumeist nicht in der Lage, Fehlerursachen zu erkennen oder zu beheben. Sechs der Befragten greifen auf ein einheitliches Notfallverfahren zurück und halten zumindest ein Trouble-Ticket-System bereit – auch wenn es in der Praxis nicht von allen Mitarbeitern angenommen und oftmals der individuelle Dienstweg vorgezogen wird. Daher vollzieht sich die Fehlerbehebung in drei Fällen auch ganz über die persönliche Kommunikation zwischen Nutzer und IT-Abteilung. Insbesondere kleine Institute begünstigen diese Form. Nichtsdestotrotz betreiben drei Institute sogar ein proaktives Monitoring, um Fehler frühzeitig erkennen zu können, ohne dass Mitarbeiter selbst aktiv werden müssen. Ein Institut bedient sich einer Support-Liste, um die Mitarbeiter zunächst zu grundlegenden Reparaturmaßnahmen ertüchtigen zu können; eine Infrastruktureinrichtung eines auf SMS basierten Meldesystems.

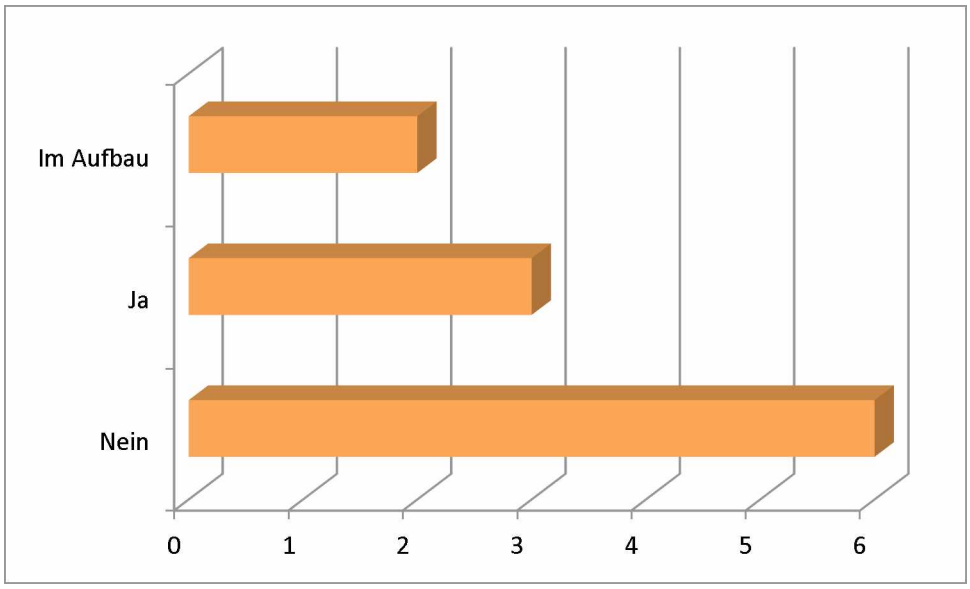
7.6 Virtualisierung der IT-Infrastruktur

Frage 49: *Spielt das Thema Virtualisierung der IT-Infrastruktur eine Rolle bei Ihren Planungen?*



Um bisweilen sehr heterogene Einzelkomponenten (Server, Netzwerke, Desktop, Software, Storage) eines Systems in einer homogenen Umgebung zusammenzufassen, wodurch sich das IT-Management deutlich verschlanken, Systeme flexibler skalieren, Prozesse deutlich transparenter gestalten und Hardware (und mithin Investitions- und Betriebskosten) einsparen lassen, werden zunehmend die Vorteile von Virtualisierungen genutzt. Alle befragten Institute und Infrastruktureinheiten führen daher eine Virtualisierung ihrer Systemkomponenten durch bzw. befinden sich gerade in dieser Umstrukturierungsphase (drei von zwölf).

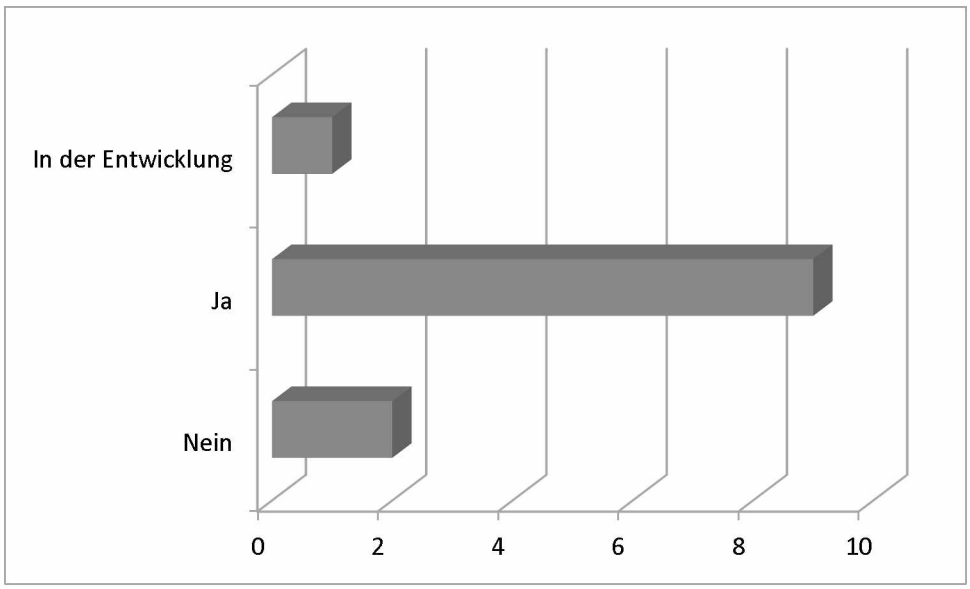
Frage 50: *Gibt es Überlegungen Grid- oder Cloudcomputing institutionell zu nutzen oder tun Sie dies bereits?*



Informationen oder Ressourcen netzbasiert und kollaborativ auszutauschen und zu bearbeiten ist die Zielsetzung von grid- und cloudbasierten Technologien. Die Lösungen reichen hierbei von Open-Source-Plattformen der gemeinsamen Datenerzeugung, -bearbeitung und -archivierung bis hin zu kommerziellen Online-Diensten zur Dokumentenbearbeitung und Ablage sowie Projektmanagementtools. Grid- und cloudbasierte Forschung findet derzeit noch keine weite Verbreitung bei den befragten Institutionen. Lediglich drei Befragte gaben an, bereits solche Ansätze zu verfolgen, wovon zwei Infrastruktureinheiten dieselbe Virtuelle Forschungsumgebung für Editionsprojekte nutzen. Zwei weitere planen die Arbeit im Grid oder in der Cloud, während sechs Häuser dies nicht tun bzw. nicht zu tun gedenken.

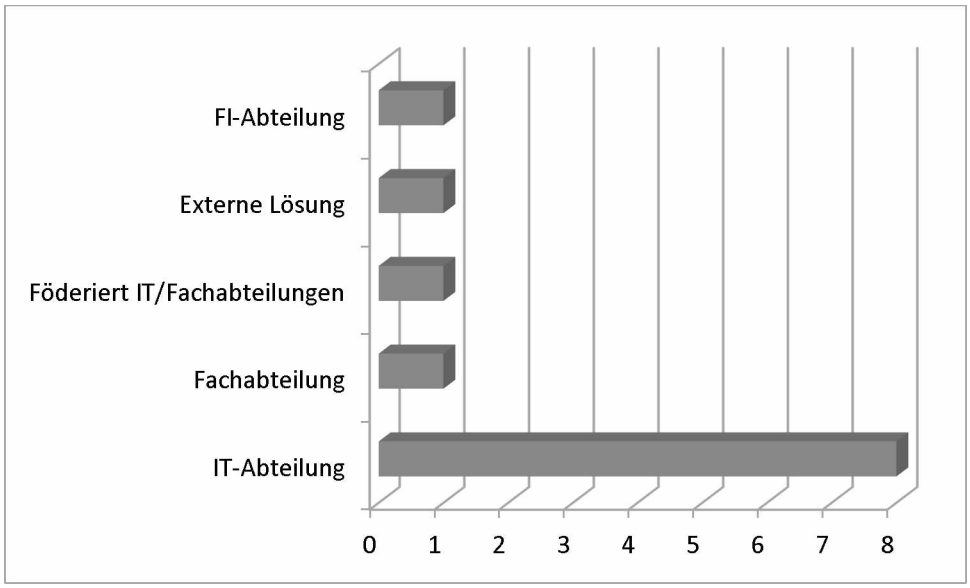
7.7 Datensicherung

Frage 51: *Gibt es eine zentrale Datensicherungsstrategie?*



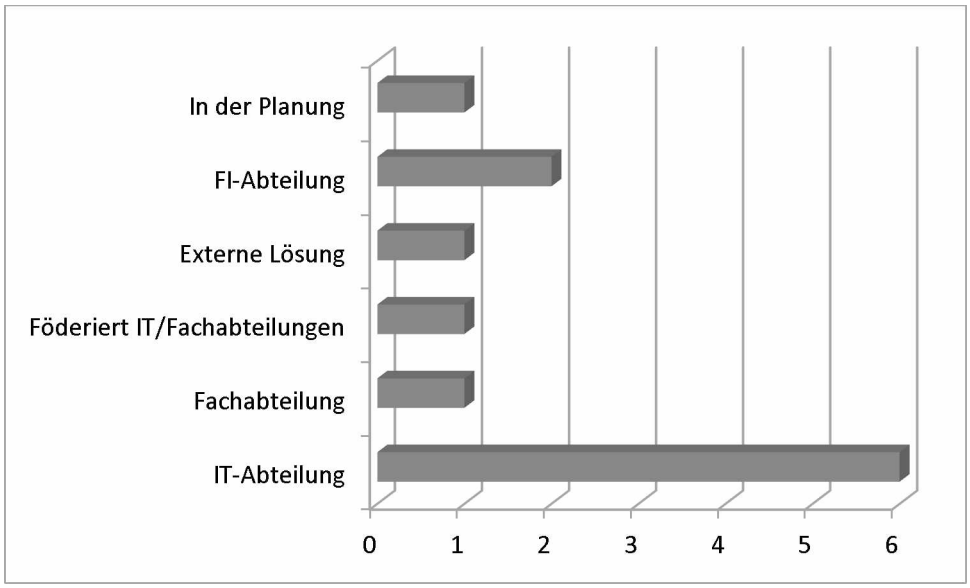
Datenverlust durch technisches oder menschliches Versagen bei unzureichend orchestrierten Insellösungen der Datensicherung entzieht wissenschaftlichen Einrichtungen die empirische Grundlage ihres Arbeitens. Ihre Wiederherstellung erfordert, wenn möglich, viel Arbeitsaufwand und ist bisweilen unzumutbar kostspielig. Aus diesem Grund haben sich die meisten Institutionen eine zentrale Datensicherungsstrategie auferlegt (neun von zwölf) oder erarbeiten diese momentan (ein Institut). Zwei Institutionen verfolgen derzeit noch individuelle Lösungen.

Frage 52: *Wer ist für die Durchführung des Backups zuständig?*



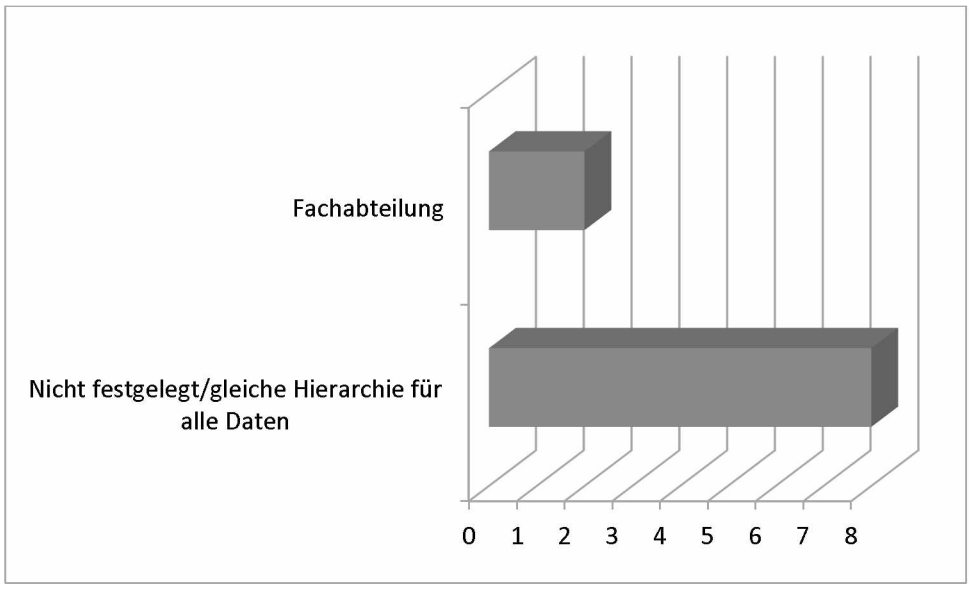
Die Zuständigkeit für die Datensicherung obliegt mehrheitlich den IT-Abteilungen (acht von zwölf). In Einzelfällen, insbesondere wenn am Institut keine eigene Organisationseinheit für IT-Fragen eingerichtet wurde, sucht man alternative Lösungen wie das Auslagern der Datensicherung oder belässt die Verantwortung ganz bei den Fachabteilungen. In einem Fall kooperieren die zentrale IT mit der jeweiligen Fachabteilung in der spezifisch angepassten und flexiblen Datensicherung; in einem anderen obliegt diese Aufgabe der zentralen FI-Abteilung.

Frage 53: *Wer bestimmt die Sicherungsintervalle?*



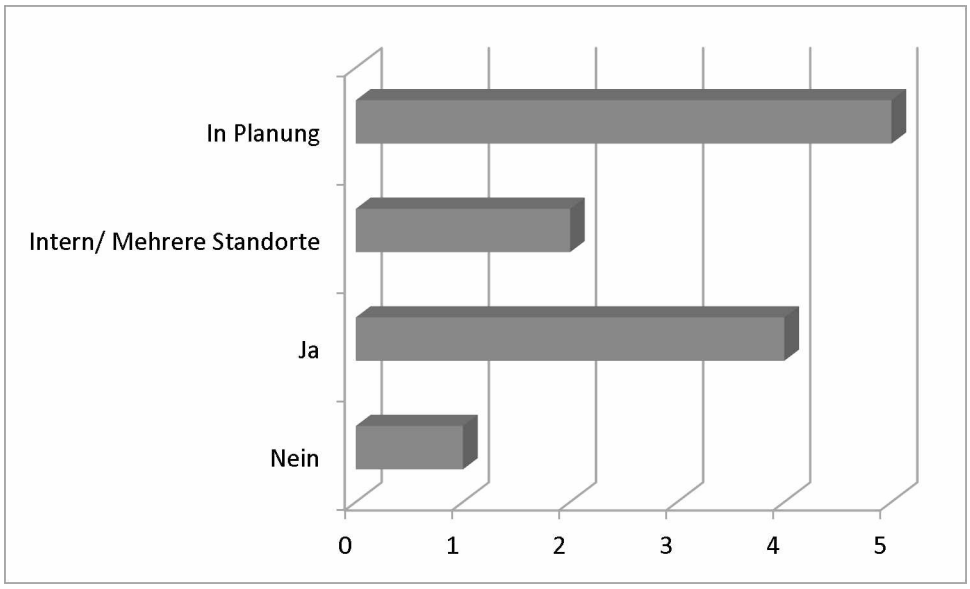
Maßgeblich für die Absicherung der Forschungsergebnisse und zur Minimierung des Aufwandes der Datenrettung sind Sicherungsintervalle, welche die Häufigkeit des Backups bestimmen. Sechs Institute und zwei Infrastruktureinrichtungen mit einer zentralen Datensicherungsstrategie beauftragen zumeist ihre IT-Abteilung mit der Definition der Sicherungsintervalle. Ein Institut, das die Datensicherung in Gemeinschaft von IT-Abteilung und Fachabteilungen durchführt (vgl. Frage 53), dezentralisiert diese Entscheidung ebenso auf diese beiden Organisationsabteilungen. Die vollverantwortlichen Fachabteilungen eines anderen Instituts planen derzeit ein abgestimmtes Vorgehen, während ein weiteres, das externe Lösungen für IT-Fragen wählt, sich der Praxis des Dienstleisters anschließt. Auffällig und ein Einzelfall ist die Arbeitsteilung eines Instituts, das die technische Ausführung des Backups seiner IT-Abteilung überlässt, die Entscheidung über die Sicherungsintervalle aber der zentralen Abteilung für Forschungsinfrastrukturen überantwortet; die dritte Forschungsinfrastruktureinrichtung fasst beide Aufgaben in ihrer FI-Abteilung zusammen.

Frage 54: *Wer legt die Sicherungswürdigkeit einzelner Objekte fest?*



Eine Datensicherung ohne Hierarchisierung kann dazu führen, dass der Datenbestand zu rasch anwächst, ohne dass die Gewährleistung getroffen wurde, dass die unterschiedslose Aufbewahrung aller gespeicherten Daten der wissenschaftlichen oder administrativen Arbeit des Hauses gerecht wird oder dem besonders essenziellen empirischen Bestand Vorrang eingeräumt würde. Dies zu verhindern ermöglicht eine Priorisierung der Daten nach Sicherungswürdigkeit. In den meisten befragten Einrichtungen ist eine solche Unterscheidung nicht festgelegt (acht von zwölf). Zwei überlassen die Entscheidung den Wissenschaftlern selbst, wobei eines dieser Institute die Zuständigkeiten in diesem Punkt ohnehin föderiert organisiert.

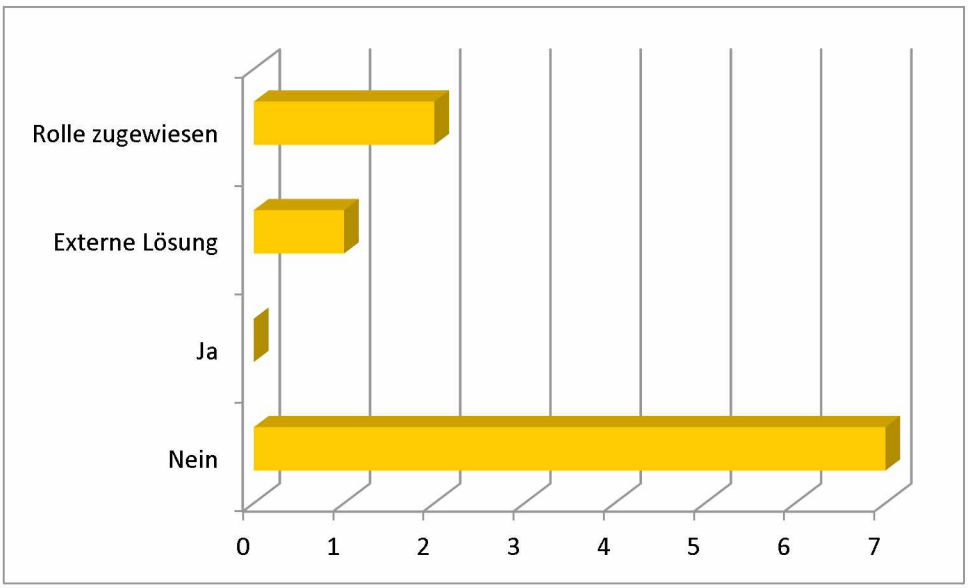
Frage 55: *Gibt es Kooperationen mit externen Partnern zur Auslagerung von Daten zum Backup?*



Sowohl die zu geringe eigene Kapazität zur Datenspeicherung als auch das Bedürfnis der redundanten Datenhaltung zur Verhinderung von physischem Datenverlust im Katastrophenfall (Brand, Wasserschaden usw.) können die Auslagerung des Backups an einen weiteren Strandort motivieren. Von den Befragten haben deswegen vier externe Lösungen gewählt, fünf weitere planen dies. Zwei Institute verfügen ohnehin über mehrere Liegenschaften, sodass eine verteilte Datensicherung auch intern möglich ist. Ein Institut verfolgt noch keinen solchen Ansatz.

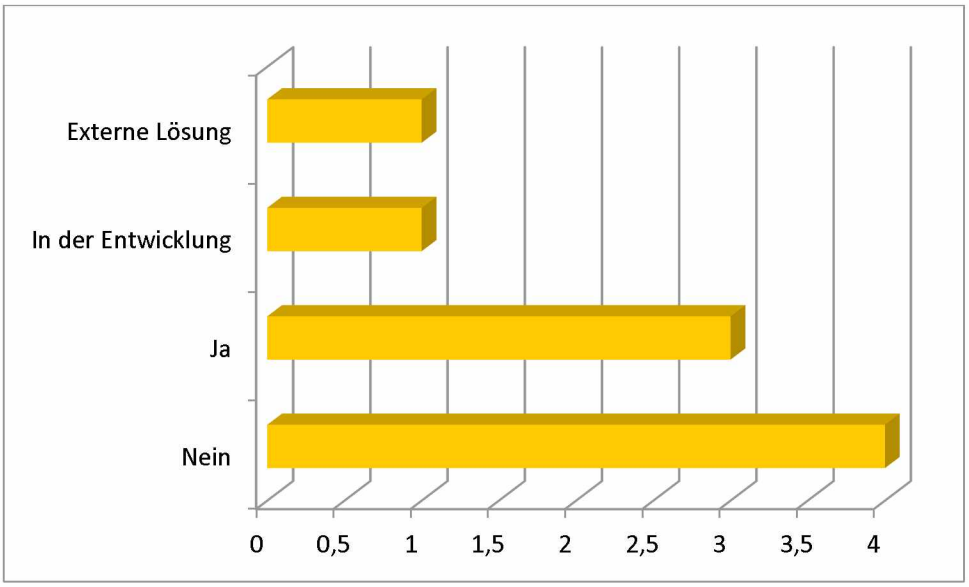
7.8 IT-Sicherheit

Frage 56: *Gibt es eine/n Sicherheitsbeauftragte/n?*



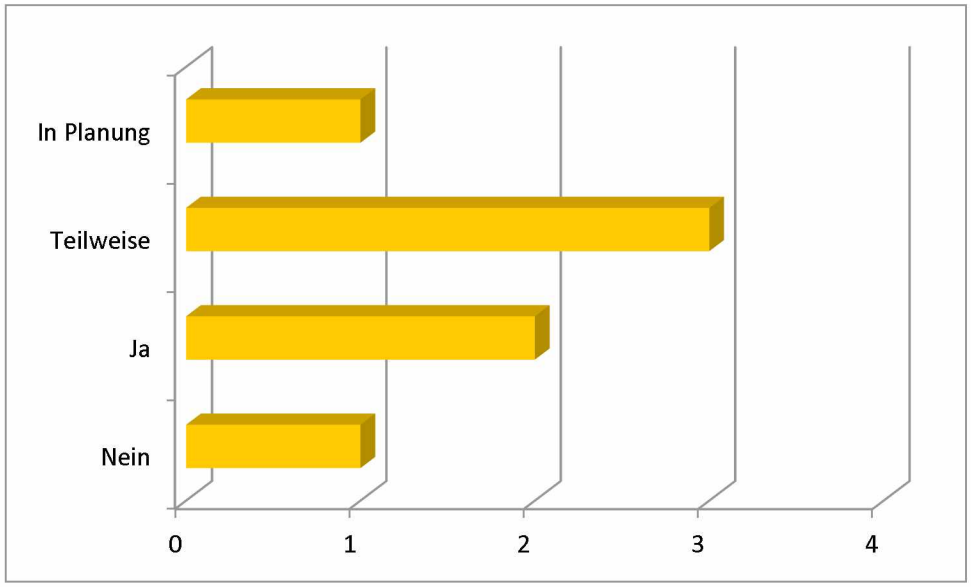
Die zunehmende Abhängigkeit der Wissenschaften von Forschungsinfrastrukturen macht sie umso anfälliger für Datenverlust oder sonstigen Schäden durch interne oder externe Sicherheitslücken. Oftmals haben die meisten Forschungsstellen noch keine dediziert für IT-Sicherheit zuständige Personalstelle ausgewiesen (sieben von zwölf). Ein Institut, das sich auch in weiteren Belangen der Forschungsinfrastrukturen auf einen externen Partner stützt, sucht auch hier die Anbindung an dessen Maßgaben. Ein anderes Institut sowie eine Infrastruktureinrichtung haben diese Funktion als Zusatzaufgabe einem Verantwortlichen im IT-Bereich zugewiesen.

Frage 57: *Gibt es institutsweite Sicherheitsrichtlinien?*



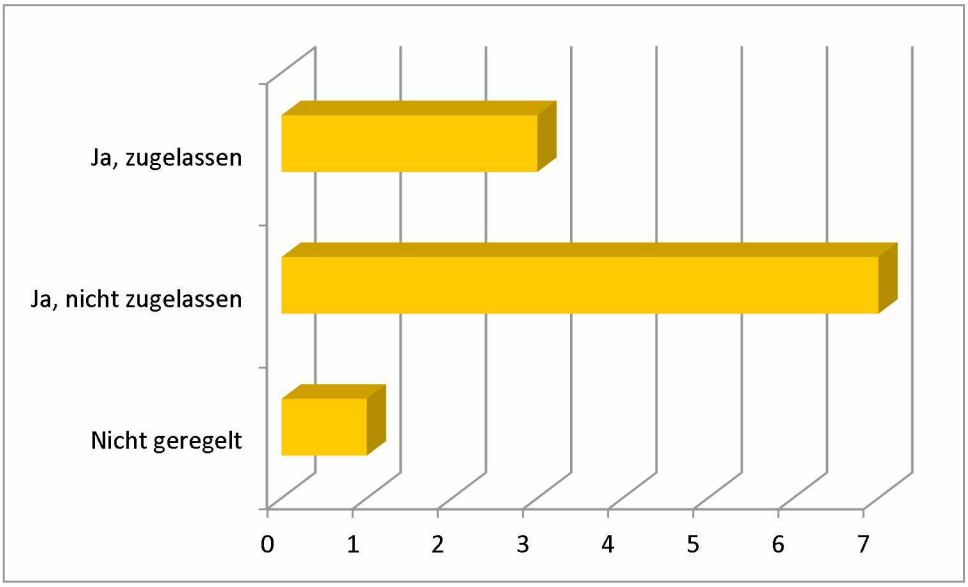
Als Gegenmaßnahme gegen interne wie externe Eingriffe in die Integrität der Systeme empfiehlt sich eine einheitliche Sicherheitsrichtlinie, die von allen Mitarbeitern verbindlich einzuhalten ist. Vier von zwölf Befragten haben ein solches Regelwerk noch nicht erarbeitet, drei haben entsprechende Maßgaben erlassen, eines entwickelt sie derzeit. Ein Institut sucht wiederum externe Lösungen (vgl. Frage 56).

Frage 58: *Haben Sie die BSI-Empfehlungen zur IT-Grundsicherung umgesetzt oder arbeiten Sie daran?*



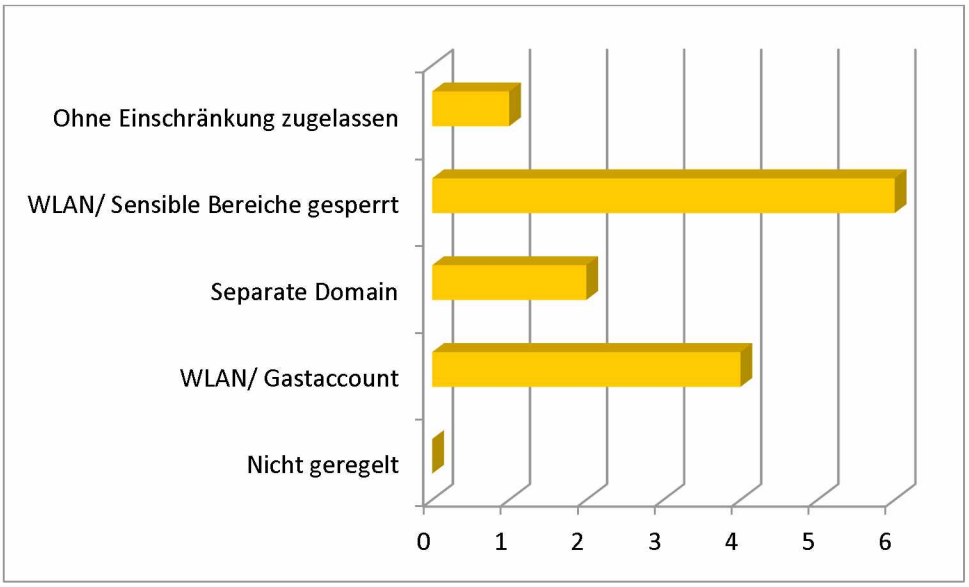
Mit dem Standard 100-2 hat das Bundesamt für Sicherheit in der Informationstechnik eine Richtlinie zum Grundschutz von IT-Infrastrukturen herausgegeben, die sich an sämtliche bundesweit tätigen Verantwortungsträger für Sicherheitsfragen im Informationsbereich wendet, um Methoden für ein effektives Sicherheitsmanagement vorzuschlagen. Vollständig wurden diese Maßgaben nur von zwei Befragten umgesetzt. Drei weitere Einrichtungen haben nur Teile der Vorschläge realisiert, da eine vollständige Umsetzung oftmals die Kapazitäten der einzelnen Häuser überschreitet bzw. manche der Maßnahmen nicht in allen Bereichen des Instituts greifen. Je ein Institut plant die Umsetzung perspektivisch, während ein weiterer Interviewpartner eigene Richtlinien erlassen hat.

Frage 59: *Ist die Nutzung mitgebrachter/privater Geräte durch Mitarbeiter und Gäste geregelt?*



Eine Kompromittierung der hausinternen Systeme (etwa durch Viren) kann oftmals, wenn auch zumeist unbeabsichtigt, durch den ungesicherten Anschluss mitgebrachter Hardware oder die Installation privater Software verursacht werden. Folglich verbieten die meisten Institute die Integration hausfremder Hardware in das interne System vollständig (sieben von zwölf). Drei lassen dies zu, während ein Institut keine Regelung getroffen hat.

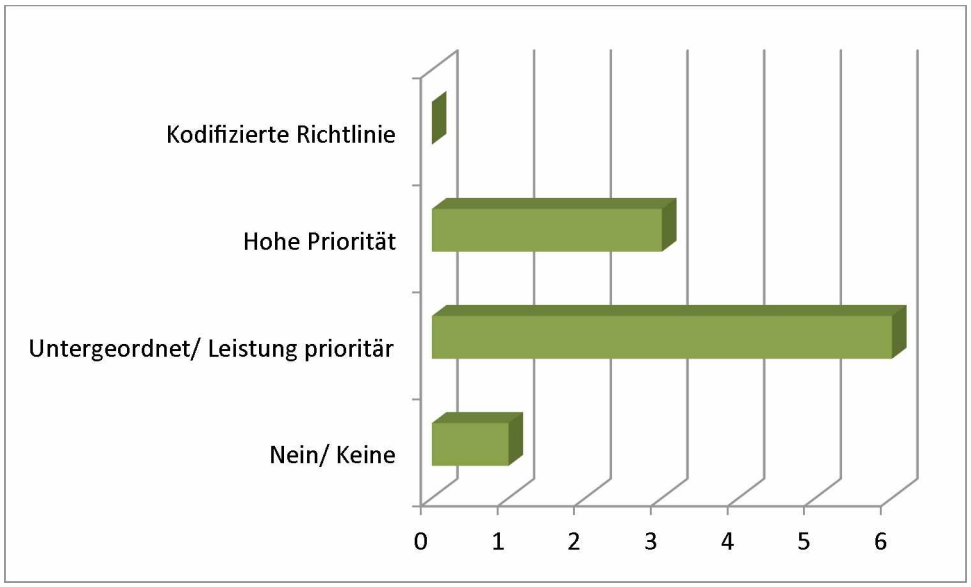
Frage 60: *Welche Strategie ist für Gäste vorgesehen?*



Die gängigste Maßnahme gegen Schäden durch nicht autorisierte Hardware erfolgt durch die Sperrung sensibler Bereiche des Netzwerks bzw. des WLANs für Gäste (sechs von zwölf). Üblicherweise wird für Nichtmitarbeiter ein gesonderter Zugang zu den Ressourcen mit eingeschränkter Funktionalität vergeben (vier von zwölf). Zwei der Befragten haben sogar eine separate Domain für Institutsgäste eingerichtet, die sie von den Ressourcen der Mitarbeiter getrennt hält. Ein Institut, das auch die Integration von fremder Hardware uneingeschränkt zulässt, gestattet auch den vollen Zugriff auf das System durch Gäste.

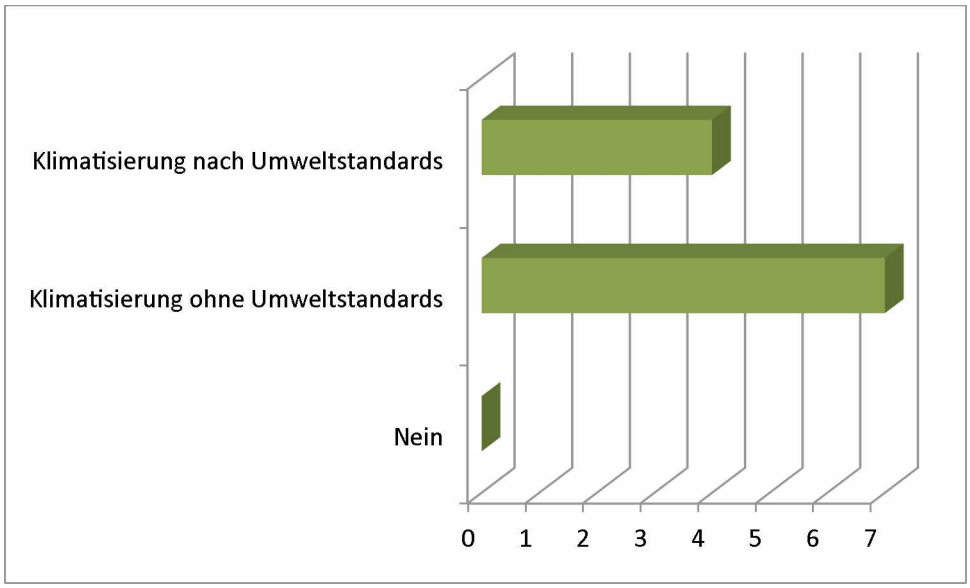
7.9 Green-IT

Frage 61: *Ist das Thema Green-IT ein Aspekt beim Betrieb oder der Anschaffung von IT-Komponenten? Welche Rolle spielt der Energieverbrauch bei der Anschaffung von Hardware?*



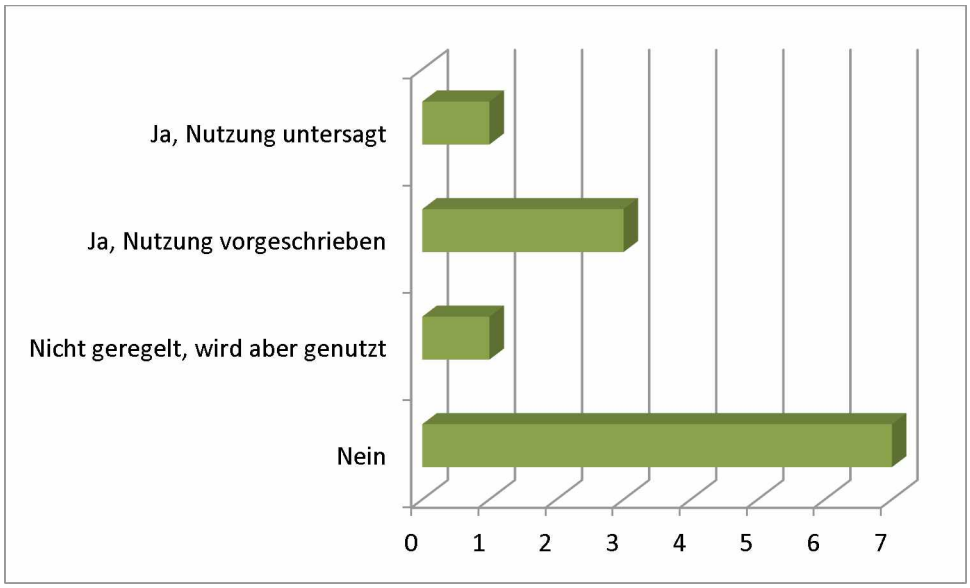
Steigende Kosten für Strom oder der zunehmende Konsens über Fragen der Umweltverträglichkeit von Arbeit und Alltag haben die Sensibilität für Green-IT in den letzten Jahren erhöht. Dennoch sehen viele der Befragten nicht zuletzt mit Blick auf eng bemessene Haushalte die zur Erfüllung ihrer wissenschaftlichen Arbeiten notwenigen Leistungsparameter als eher prioritär gegenüber Umweltfragen an (sechs von zwölf). Drei Interviewte messen diesem Themenkomplex bei ihren Planungen eine gewisse, ein Institut hingegen keine Relevanz bei. Nichtsdestotrotz hat keiner der Befragten eine umfassend kodifizierte Richtlinie zum Umweltschutz erlassen.

Frage 62: *Gibt es im Rechenzentrum eine moderne Klimatisierung?*



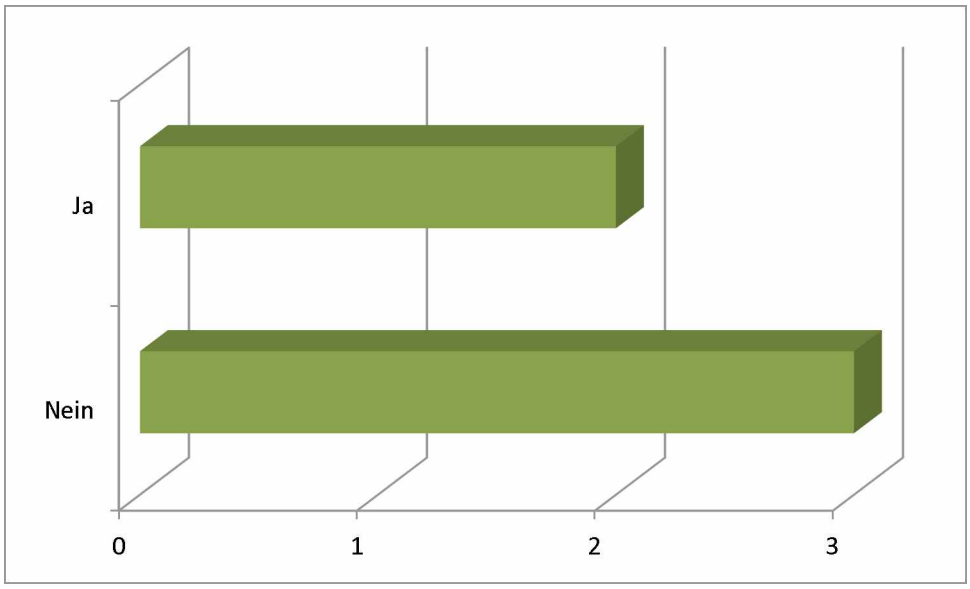
Kostenintensiv im Betrieb von Forschungsinfrastrukturen gestaltet sich zumeist die Klimatisierung des Rechenzentrums. Neue Technologien wie Blindblenden, Kabeldurchführungen und insbesondere Einhausungen ermöglichen eine deutliche Reduzierung des Stromverbrauchs. Vier von zwölf Instituten haben bereits Maßnahmen zur kostensensibleren Planung ihres Serverraums ergriffen, während sieben Häuser noch über konventionelle Klimatechnik verfügen.

Frage 63: *Gibt es Vorschriften zur Nutzung von Umweltpapier?*



Hohe Mengen an Bedrucktem gehören, oft auch als Abfallprodukt, zu den ressourcenintensiven Erzeugnissen des wissenschaftlichen Arbeitens mit Forschungsinfrastrukturen. Die Verwendung von Umweltpapier findet daher insbesondere seit der deutlichen Verbesserung von Optik und Qualität deutlich mehr Zustimmung. Von zwölf Befragten statten sieben ihre Drucker und Kopierer nicht mit Umweltpapier aus. Drei Institute haben eine Vorschrift für die Nutzung erlassen, während ein Institut Recyclingfähiges Papier eher auf freiwilliger Basis der verantwortlichen benutzt. Ein Institut hat hingegen ein explizites Verbot dieses Papiers ausgesprochen.

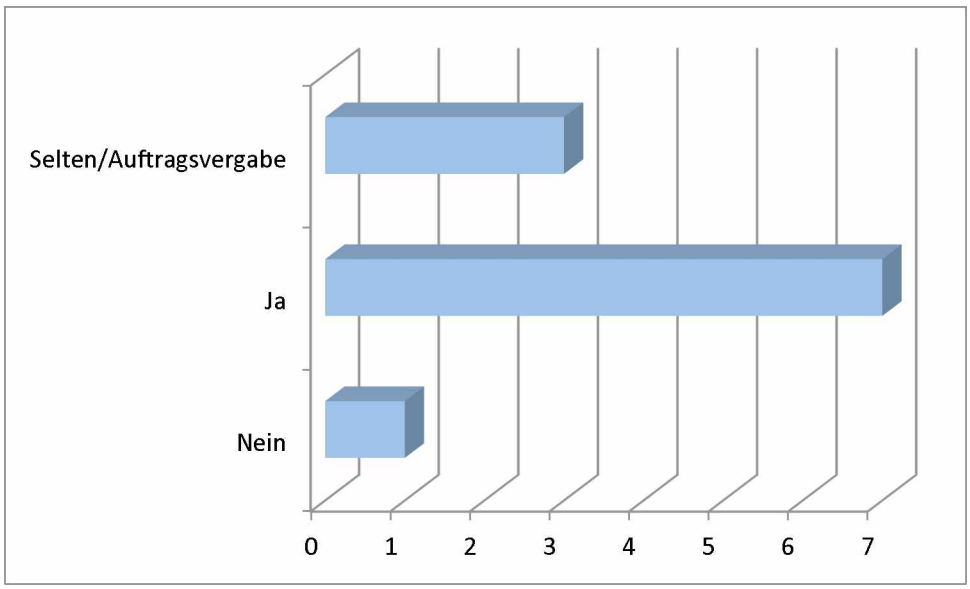
Frage 64: *Wenn Sie sich noch nicht explizit mit dem Thema Green-IT befasst haben, planen Sie dies eventuell für die Zukunft?*



Von den vier Instituten ohne verbindliche Vorgaben zu Green-IT gedenken zwei künftig eine höhere Priorisierung stromsparender Hardware vorzunehmen. Die übrigen drei planen jedoch keine Änderung der bisherigen Abstinenz.

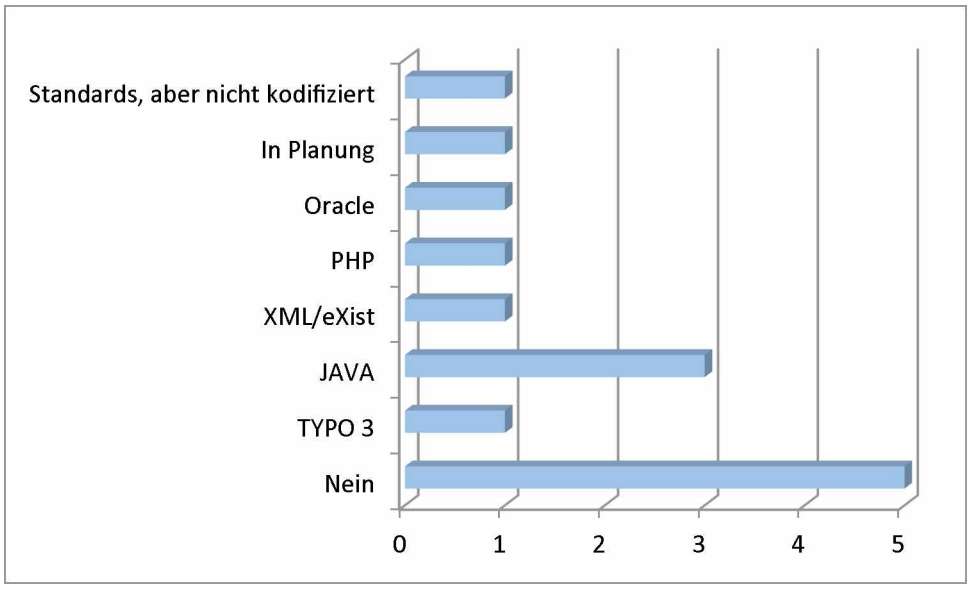
7.10 Software-Entwicklung

Frage 65: *Gibt es an Ihrem Institut Spezialsoftware, die Sie selbst entwickeln?*



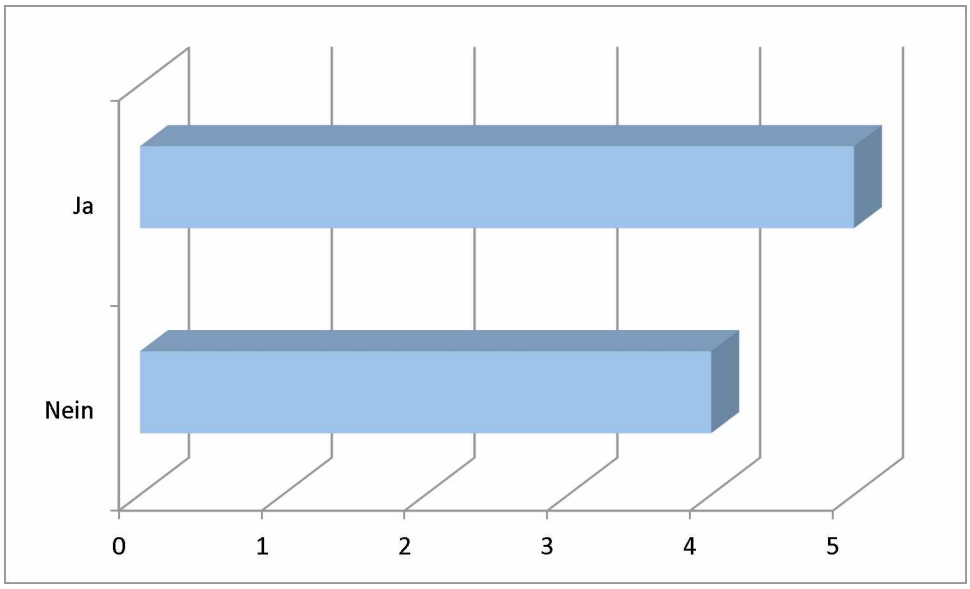
Aufgrund teilweise hochspezialisierter Forschungsfragen und Methoden bieten viele frei erhältliche Softwareprodukte nicht die notwendigen Funktionalitäten, um für Forschungsinstitute vollumfänglich tauglich zu sein. Daher wird gelegentlich auf Programmierfähigkeiten der eigenen Mitarbeiter zurückgegriffen. So haben auch sieben der zwölf Befragten angegeben, eigene Software häufig entwickelt zu haben. Drei weitere Institute tun dies selten und lediglich im Projektrahmen im Verbund mit Partnern oder vergeben Neuentwicklungen an externe Dienstleister. Ein Institut hat keine Eigenentwicklungen vorgenommen.

Frage 66: *Gibt es institutsweite Richtlinien zur Nutzung bestimmter Programmiersprachen oder -umgebungen oder ist dies eine persönliche Entscheidung des Entwicklers?*



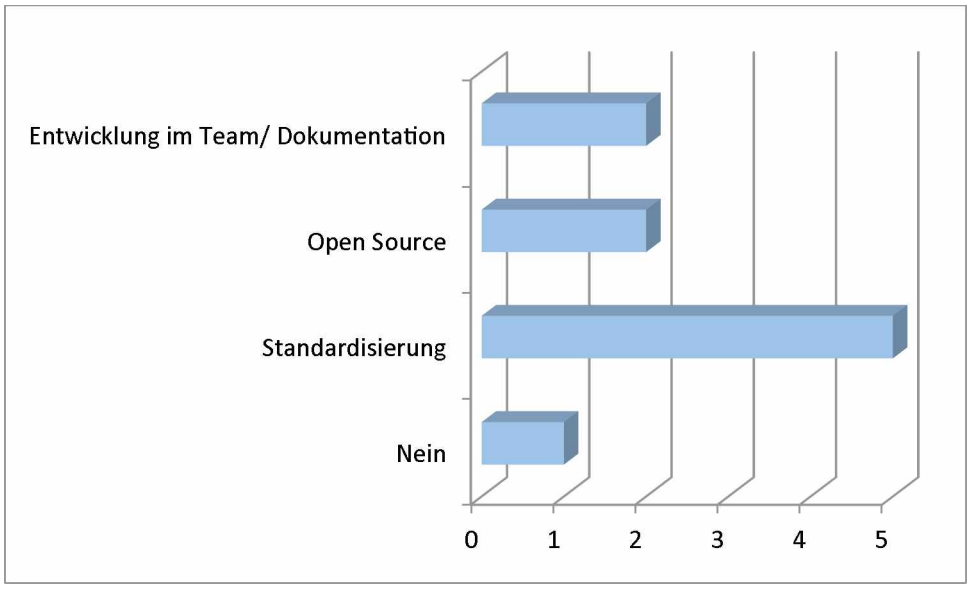
Bei der Erstellung eigener Algorithmen bis hin zu speziellen Tools oder Softwareprodukten kann im Sinne der Nachnutzbarkeit, Nachhaltigkeit und Kompatibilität die Verwendung gängiger und einheitlicher Programmiersprachen und Umgebungen nützlich sein. Unter den Befragten fanden sich fünf, die verbindliche Vorgaben zumindest planen oder mit TYPO3, für den Internetauftritt des Hauses, JAVA, XML oder PHP bereits einheitliche Standards verwenden. Fünf der Interviewten entwickelt entweder keine Software oder verpflichtet seine Mitarbeiter diesbezüglich nicht zur Einhaltung von Standards.

Frage 67: *Gibt es bestimmte Standardsoftware, auf denen eigene Entwicklungen aufgebaut werden?*



Analog hierzu besteht bei der Entwicklung eigener Anwendung die Möglichkeit der Verwendung von Standardsoftware als Grundlage weiterführender Produkte. Vier der Institute verneinten die Frage nach verbindlichen Regelungen in diesem Punkt, fünf weitere haben entsprechende Maßgaben erlassen.

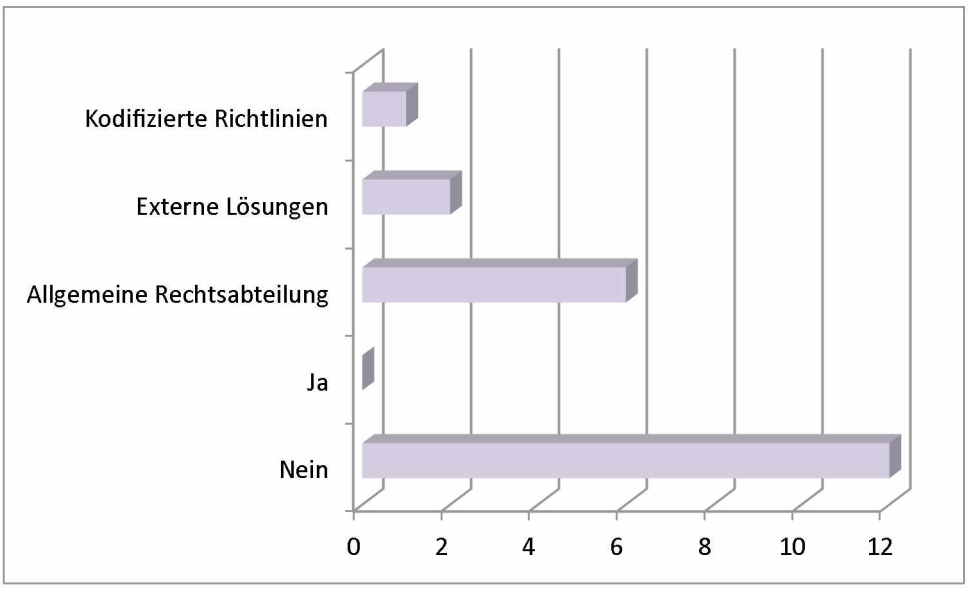
Frage 68: *Gibt es Lösungen zur Verstetigung von Eigenentwicklungen?*



Eigenentwicklungen entsprechen zwar häufig spezifischen Anforderungen der Wissenschaft besser als kommerzielle Produkte, mangeln aber aufgrund tendenziell proprietärer und nichtstandardisierter Auslegung an nachhaltiger Anwendbarkeit. Fünf der befragten Einrichtungen streben deswegen zumindest innerhalb des Hauses standardisierte Grundlagen für Eigenentwicklungen an. Die Bereitstellung als Open-Source-Produkt erlaubt es einer wissenschaftlichen Öffentlichkeit im Falle zweier weiterer Institute eine gemeinsame und nachhaltige Softwarepflege. Zwei Institute verlassen sich auf redundante Personalstrukturen bei der Entwicklung und eine sorgfältige Dokumentation des Quellcodes.

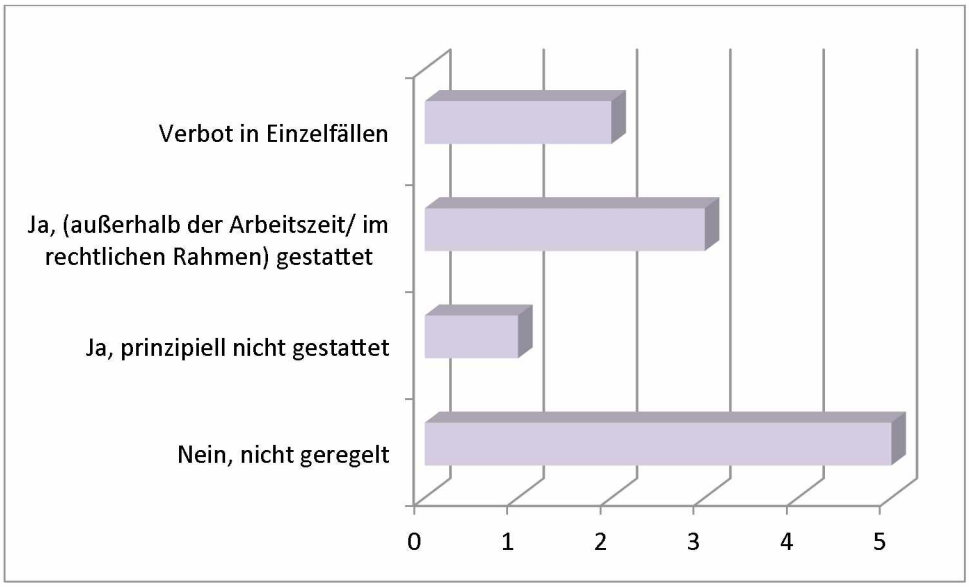
7.11 Juristische Aspekte

Frage 69: *Gibt es eine Rechtsabteilung mit expliziten Zuständigkeiten für IT-Fragen?*



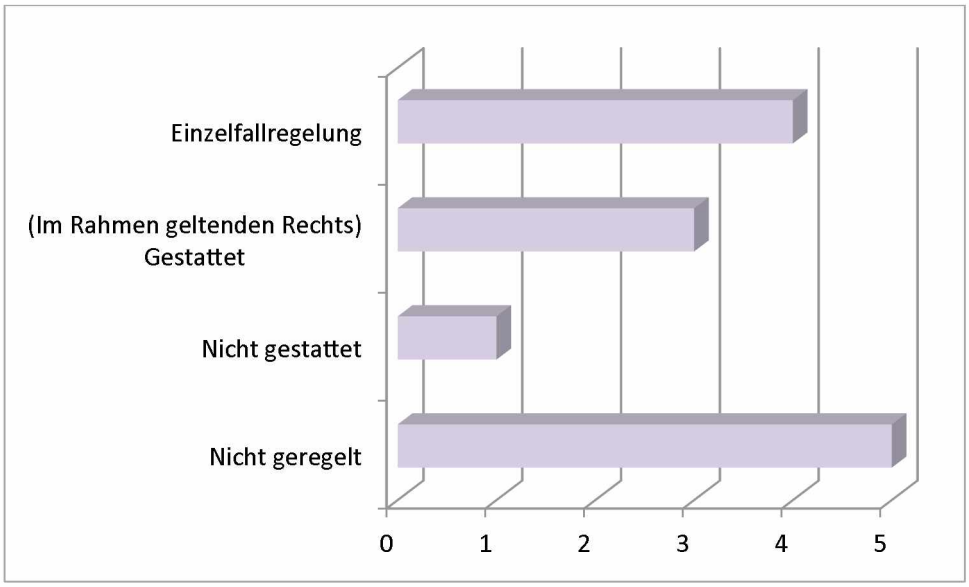
Die Nutzung von Infrastrukturkomponenten und Forschungsdaten kann eine ganze Reihe von juristischen Herausforderungen aufwerfen, die das Urheber- oder Datenschutzrecht betreffen. Keiner der Befragten hat – oft aus Kapazitätsgründen – eine dediziert auf solche Fragen spezialisierte Rechtsabteilung eingerichtet. Indes verfügen sechs Häuser über eine eigene allgemein zuständige Rechtsabteilung oder einen Justiziar, die im Zweifelsfalle tätig werden müssten. Zwei Institute greifen in Rechtsfällen auf externe Expertise zurück, wovon ein Institut intern explizite juristische Richtlinien für den Umgang mit der eigenen Forschungsinfrastruktur ausgearbeitet hat.

Frage 70: *Gibt es eine Betriebsvereinbarung zur privaten Internetnutzung?*



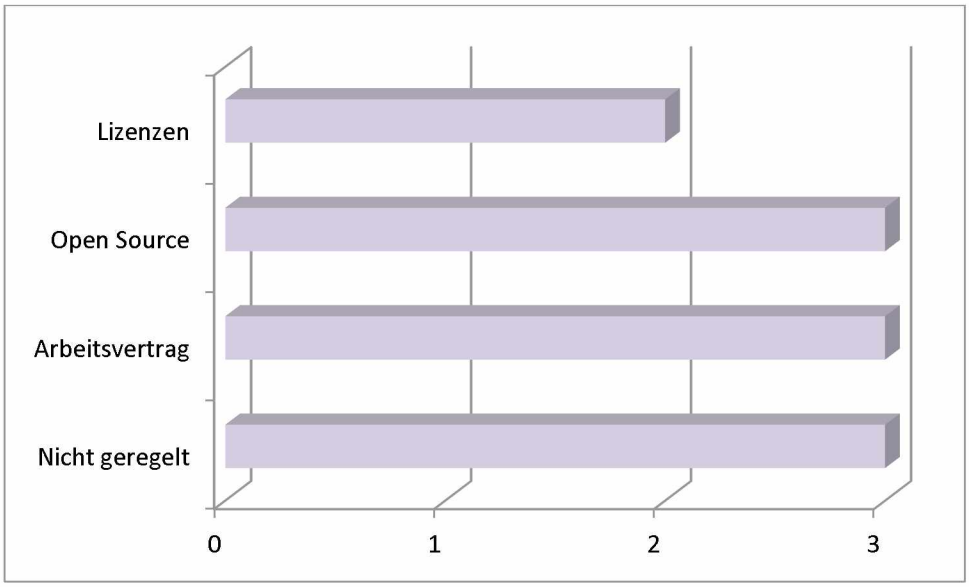
Betriebsvereinbarungen sind das gängige Medium, um das institutsinterne Zusammenspiel von Mitarbeitern und Ressourcen zu regeln. Hierbei ist der tagtägliche Umgang mit dem Internet zu dienstlichen und privaten Zwecken auf Einrichtungen des Instituts eine der häufigsten Fragen, die der Regulierung bedürfen. Fünf Institute haben diesen Punkt in ihrem internen Regelwerk nicht explizit kodifiziert. Drei der Befragten gestatten eine private Nutzung außerhalb der Dienstzeit im Rahmen geltenden Rechts, während zwei weitere Häuser zwar in begründeten Einzelfällen Verbote ausgesprochen, aber keine allgemeingültige Regelung erlassen haben. Nur einer der Interviewten Partner untersagt eine außerdienstliche Nutzung explizit.

Frage 71: Gibt *es Regelungen zur Cloud-Nutzung oder zur Nutzung von Skype?*



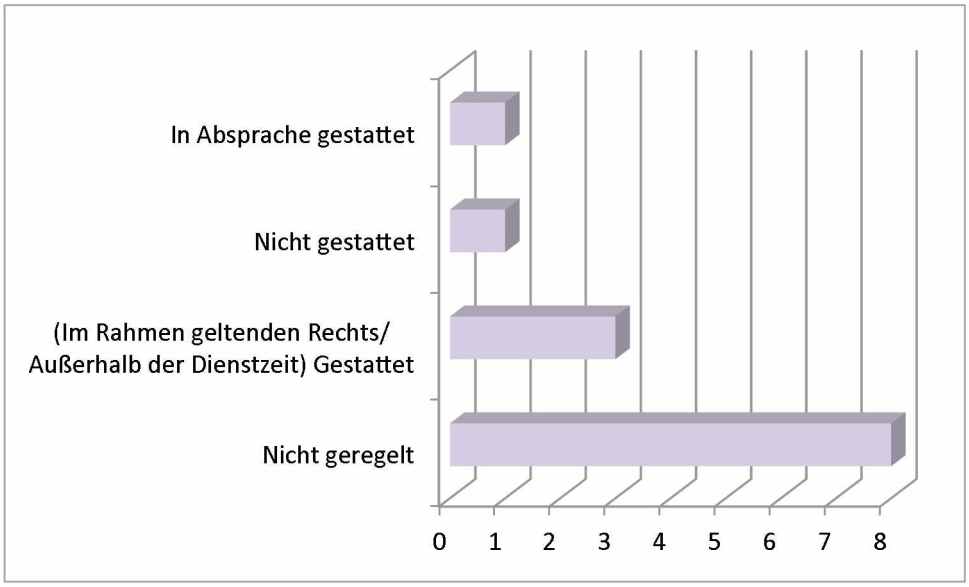
Ungeachtet des koordinatorischen Vorteils der Nutzung von Cloud-Diensten wie Skype oder Google Docs vor allem für dislozierte Verbundprojekte können datenschutzrechtliche Bedenken hinsichtlich des Austausches von Forschungsdaten und Institutsinterna über solche Dienste entstehen. Dies gilt insbesondere dann, wenn Nutzer dieser Dienste keinen Einfluss auf sich ändernde Nutzungsbestimmungen haben. Bei fünf der Befragten existieren keine Richtlinien zum Umgang mit diesen Diensten; drei gestatten die Nutzung im Rahmen geltenden Rechts. Von den Instituten, die rechtliche Probleme sehen, verbietet eines die Nutzung vollständig, während sie vier durch fehlende technische Unterstützung oder speziellen Empfehlungen im Einzelfall verhindern.

Frage 72: *Sind Urheberrechte bzw. Verwertungsrechte geregelt?*



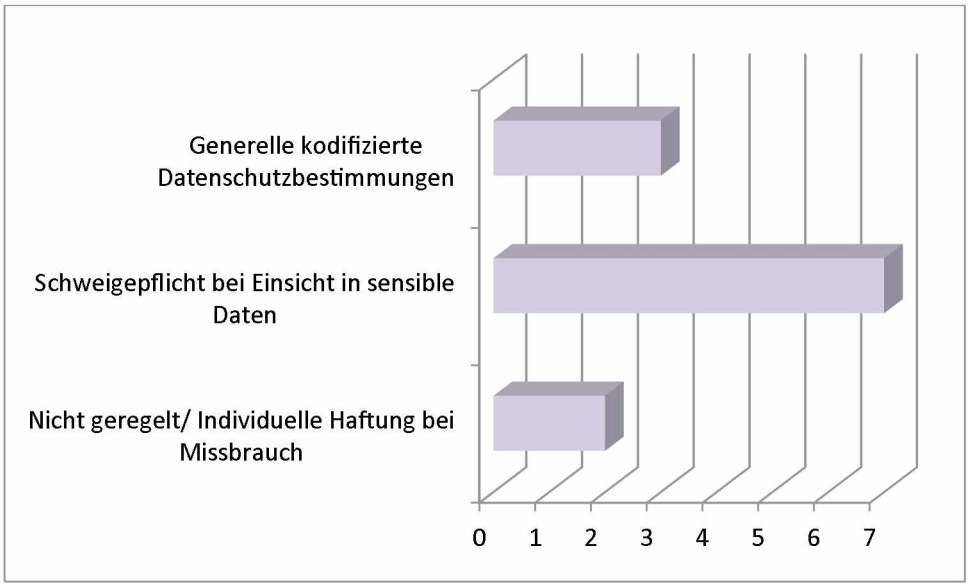
Bei der Generierung von Forschungsdaten und der Erstellung eigener Software können sich urheberrechtliche Fragen aufwerfen – gilt doch der einzelne Wissenschaftler und nicht seine Forschungsinstitution als Urheber seiner Erzeugnisse, hält der Arbeitgeber in der Regel die Verwertungsrechte. Gleichzeitig können Forschungsdaten und Erkenntnisse an lizenzrechtliche Beschränkungen gebunden sein, die eine Weitergabe jenseits der Vertragsparteien oder eine kommerzielle Nutzung ausschließen. Dies über entsprechende Verträge – inklusive der Arbeitsverträge der Mitarbeiter – zu regeln, liegt somit im Interesse der Arbeitsgeber. Drei der Befragten übertragen Urheberrechte für Forschungsergebnisse im Rahmen von Arbeitsverträgen vom Individuum auf die Forschungsinstitution. Drei weitere Einrichtungen stellen alle Ergebnisse grundsätzlich als Open Source bzw. Open Access zur Verfügung, sodass sich rechtliche Schranken gar nicht erst aufbauen. Eine Wissensgenerierung, die dies nicht zuließe, würde nicht stattfinden. Zwei Interviewpartner lizenzieren ihre Ergebnisse und steuern die Nachnutzung somit selbst; alle Infrastruktureinrichtungen erlassen indes keine Vorgaben.

Frage 73: *Gibt es Regelungen zur Nutzung der institutseigenen IT-Infrastruktur für private Arbeiten?*



Neben dem Kernauftrag des einzelnen Wissenschaftlers, der ihn seinem Arbeitgeber verpflichtet, prägen berufsnahe Aktivitäten wie die Betreuung von Studierenden, akademische Lehre, Eigenpublikationen oder die Ausarbeitung eigener Abschlussarbeiten den wissenschaftlichen Alltag. Unabhängig von dem Nutzen, den ein Institut aus solchen Aktivitäten ziehen könnte, stellt sich für jeden Arbeitsgeber die Frage, ob dieser Ressourcen hierfür bereitzustellen gewillt ist. Die überwiegende Zahl der Befragten (acht von zwölf) hat hierzu keine Regelungen erlassen. Drei Institute gestatten solche semi-private Tätigkeiten außerhalb der Dienstzeit, sofern geltendes Recht beachtet wird. Lediglich ein Institut untersagt sie vollständig.

Frage 74: *Wie sind datenschutzrechtliche Verpflichtungen an Ihrem Institut geregelt?*



Insbesondere Systemadministratoren unterliegen selten datenschutzrechtlichen Beschränkungen wie sie für wissenschaftliche Mitarbeiter oft gelten. Es empfiehlt sich somit, diesem in teils hochsensiblen Bereichen tätigen Personal gesonderte Verschwiegenheitspflichten aufzuerlegen. Sieben Interviewpartner haben bereits entsprechende Maßnahmen umgesetzt, während einschlägige Regelungen in drei weiteren im Rahmen der allgemeinen Datenschutzbestimmung des Hauses Eingang gefunden haben. Schließlich verweisen die zwei verbleibenden Einrichtungen auf die individuelle Haftung aller Mitarbeiter bei jeweils geltendem Recht, ohne eigene Richtlinien herauszugeben.

8. Zusammenfassung, Diskussion, Schlussfolgerung

Digitale Forschungsinfrastrukturen sind auch in außeruniversitären Forschungseinrichtungen ein wesentlicher Bestandteil. Die Aufgaben, die durch die beteiligten Institutionen wie zentraler Datenverarbeitung und Bibliothek zu bewältigen sind, sind umfangreich und erfordern profunde Kenntnisse sowohl in Bezug auf die Technik als auch auf juristische Fragen. Kleinere Forschungsinstitutionen können hier von einer Politik der kurzen Wege profitieren, die eine Kooperation innerhalb der Einrichtung zwischen den Akteuren wie Bibliothek und ZDV erleichtert. Darüber hinaus empfiehlt sich für größere Forschungsinfrastrukturen wie die Einrichtung eines Dokumentenservers oder einer langfristigen Backup- und Archivierungsstrategie die Zusammenarbeit in größeren Verbünden, um Know-How und finanzielles Engagement über Standorte zu verteilen. Auch sind Spezialisierungen einzelner Einrichtungen auf spezifische Fragen (technische Fragen der Archivierung, Zugangssysteme, IT-Infrastruktur) denkbar und sinnvoll.

9. Literatur

Bargheer, M., Bellem, S. & Schmidt, B. (2006) Open Access und Institutional Repositories – Rechtliche Rahmenbedingungen. In G. Spindler, ed. *Rechtliche Rahmenbedingungen von Open Access-Publikationen*. Göttinger Schriften zur Internetforschung. Göttingen: Universitätsverlag Göttingen, S. 1–20. URL: http://www.univerlag.uni-goettingen.de/OA-Leitfaden/oaleitfaden_web.pdf [Zuletzt abgerufen am 9. April, 2014].

Bertelmann, R. (2006) Vom Dokumentenserver zum Institutional Repository. In H.-C. Hobohm & K. Umlauf, eds. *Erfolgreiches Management von Bibliotheken und Informationseinrichtungen*. Hamburg: Dashöfer.

Borel, F., Lienhard, J., Oberknapp, B., Ruppert, A. & Steilen, G. (2009) Authentifizierung und Autorisierung mit Shibboleth in der Föderation DFN-AAI. *b.i.t.online* 12, S. 285–288.

Bradford, E. & Mauget, L. (2002) *Linux and Windows Interoperability Guide*, Prentice Hall Professional.

Bräuninger, M., Haucap, J., Stepping, K. & Stühmeier, T. (2012) *Cloud Computing als Instrument für effiziente IT-Lösungen*, Hamburg: Hamburgisches WeltWirtschaftsInstitut.

Bundesamt für Sicherheit in der Informationstechnik (2008a) *IT-Grundschutz-Vorgehensweise. Version 2.0*, Bonn: BSI. URL: https://www.bsi.bund.de/SharedDocs/Downloads/DE/BSI/Publikationen/ITGrundschutzstandards/standard_1002_pdf.pdf?__blob=publicationFile [Zuletzt abgerufen am 9. April, 2014].

Bundesamt für Sicherheit in der Informationstechnik (2008b) *Notfallmanagement*, Bonn: BSI. URL: https://www.bsi.bund.de/SharedDocs/Downloads/DE/BSI/Publikationen/ITGrundschutzstandards/standard_1004_pdf.pdf?__blob=publicationFile [Zuletzt abgerufen am 9. April, 2014].

Bundesamt für Sicherheit in der Informationstechnik (2009) *Drahtlose Kommunikationssysteme und ihre Sicherheitsaspekte*, Bonn: BSI. URL: https://www.bsi.bund.de/SharedDocs/Downloads/DE/BSI/Publikationen/Broschueren/Drahtlose-Komsysteme_pdf.pdf?__blob=publicationFile [Zuletzt abgerufen am 9. April, 2014].

Bundesamt für Sicherheit in der Informationstechnik (2012) *Sicherheitsempfehlungen für Cloud Computing Anbieter (Mindestanforderungen in der Informationssicherheit)*, Bonn: BSI. URL: https://www.bsi.bund.de/SharedDocs/Downloads/DE/BSI/Mindestanforderungen/Eckpunktepapier-Sicherheitsempfehlungen-CloudComputing-Anbieter.pdf?__blob=publicationFile [Zuletzt abgerufen am 9. April, 2014].

Bundesamt für Sicherheit in der Informationstechnik (2014) *Kryptographische Verfahren: Empfehlungen und Schlüssellängen*, Bonn: BSI. URL: https://www.bsi.bund.de/SharedDocs/Downloads/DE/BSI/Publikationen/TechnischeRichtlinien/TR02102/BSI-TR-02102_pdf.pdf?__blob=publicationFile [Zuletzt abgerufen am 9. April, 2014].

Bundesministerium für Bildung und Forschung (2013) *Forschungsinfrastrukturen für die Geistes- und Sozialwissenschaften*, Bonn. URL: http://www.bmbf.de/pub/forschungsinfrastrukturen_geistes_und_sozialwissenschaften.pdf.

Bundesministeriums für Bildung und Forschung (2013) *Roadmap für Forschungsinfrastrukturen. Pilotprojekt des BMBF*, Bonn. URL: http://www.bmbf.de/pub/Roadmap.pdf.

Burnard, L. & Bauman, S. eds. (2014) *TEI P5: Guidelines for Electronic Text Encoding and Interchange*, Charlottesville, Virginia: Text Encoding Initiative Consortium. URL: http://www.tei-c.org/release/doc/tei-p5-doc/en/Guidelines.pdf [Zuletzt abgerufen am 9. April, 2014].

Crow, R. (2002) *The Case for Institutional Repositories: A SPARC Position Paper*, Washington, DC: Association of Research Libraries (ARL).

Degkwitz, A. (2007) Open Access und die Novellierung des deutschen Urheberrechts. *Zeitschrift für Bibliothekswesen und Bibliographie* 54, S. 243–245.

Deloitte (2012) Flexible Working 2012: Wie flexibel gestalten Unternehmen in Österreich die Arbeit ihrer MitarbeiterInnen?

DFG (2009) *Empfehlungen zur gesicherten Aufbewahrung und Bereitstellung digitaler Forschungsprimärdaten*, Bonn: Deutsche Forschungsgemeinschaft. URL: http://www.dfg.de/download/pdf/foerderung/programme/lis/ua_inf_empfehlungen_200901.pdf [Zuletzt abgerufen am 22. April, 2014].

DFG (2010) *Informationsverarbeitung an Hochschulen – Organisation, Dienste und Systeme: Empfehlungen der Kommission für IT-Infrastruktur für 2011–2015*, Bonn: Deutsche Forschungsgemeinschaft. URL: http://www.dfg.de/download/pdf/foerderung/programme/wgi/empfehlungen_kfr_2011_2015.pdf [Zuletzt abgerufen am 22. April, 2014].

DFN-AAI (2010a) DFN-AAI – Authentifizierungs- und Autorisierungs-Infrastruktur im DFN. URL: https://www.aai.dfn.de/aktuelles/ [Zuletzt abgerufen am 28. May, 2013].

DFN-AAI (2010b) Technische und organisatorische Voraussetzungen an das Identity Managment. URL: https://www.aai.dfn.de/der-dienst/identitymanagement/ [Zuletzt abgerufen am 9. April, 2014].

Die Landesbeauftragte für Datenschutz und Informationsfreiheit (2013) Orientierungshilfe „Soziale Netzwerke". In *Konferenz der Datenschutzbeauftragten des Bundes und der Länder.*

Doyle, S. (2000) *Understanding Information Technology*, Cheltenham: Nelson Thornes.

Eisentraut, P. & Helmle, B. (2013) *PostgreSQL-Administration*, Beijing; Köln: O'Reilly Germany.

Foster, N. F. & Gibbons, S. (2005) Understanding Faculty to Improve Content Recruitment for Institutional Repositories. *D-Lib Magazine* 11. URL: http://www.dlib.org/dlib/january05/foster/01foster.html [Zuletzt abgerufen am 1. August, 2013].

Fournier, J. (2005) *Wege zum Wissen. Aktionsfelder zur Förderung des Open Access durch die DFG*, Deutsche Forschungsgemeinschaft. URL: http://dfg.de/download/pdf/dfg_im_profil/evaluation_statistik/programm_evaluation/studie_publikationsstrategien_stellungnahme.pdf [Zuletzt abgerufen am 9. April, 2014].

Gausemeier, J., Plass, C. & Wenzelmann, C. (2009) *Zukunftsorientierte Unternehmensgestaltung: Strategien, Geschäftsprozesse und IT-Systeme für die Produktion von morgen*, München [u.a.]: Hanser Verlag.

Gietz, P. (2004) Identity Management an deutschen Hochschulen. In *18. DFN-Arbeitstagung über Kommunikationsnetze.* Düsseldorf. URL: http://subs.emis.de/LNI/Proceedings/Proceedings55/GI-Proceedings.55-31.pdf [Zuletzt abgerufen am 9. April, 2014].

Gietz, P., Grimm, C., Pfeiffenberger, H., Rauschenbach, J. & Schröder, R. (2006) Auf dem Weg zur DFN-AAI: Identity Management. *DFN Mitteilungen 71*, S. 12–14.

Görl, S., Puhl, J. & Thaller, M. (2011) *Empfehlungen für die weitere Entwicklung der Wissenschaftlichen Informationsversorgung des Landes NRW*, Berlin: epubli GmbH.

Henrich, A. (2011) Forschungsinfrastrukturen - Auf dem (langen) Weg zu den e-Humanities. *uni.vers, Magazin der Otto-Friedrich-Universität Bamberg*, S. 11–15.

Hoeren, T. (2013) Internetrecht. URL: http://www.uni-muenster.de/Jura.itm/hoeren/itm/wp-content/uploads/Skript-Internetrecht_Oktober2013.pdf [Zuletzt abgerufen am 5. April, 2014].

Hohoff, U. (2011) Bessere Infrastrukturen für die geistes- und sozialwissenschaftliche Forschung. Der Wissenschaftsrat zieht Bilanz und fordert mehr Aufbauarbeit. *ABI Technik: Zeitschrift für Automation, Bau und Technik im Archiv-, Bibliotheks- und Informationswesen* 31, S. 2–10.

Ide, N. M., Bonhomme, P. & Romary, L. (2000) XCES: An XML-based Encoding Standard for Linguistic Corpora. In *Proceedings of the Second International Language Resources and Evaluation (LREC 2000).* Athen: European Language Resources Association (ELRA), S. 825–830.

ISO/IEC JTC 1/SC 27 (2005) *ISO/IEC 27002:2005: Information technology -- Security techniques -- Code of practice for information security management*, Genf.

ISO/TC 37/SC 4 (2012) *ISO 24612:2012: Language resource management -- Linguistic annotation framework (LAF)*, Genf.

Jensen, U., Katsanidou, A. & Zenk-Möltgen, W. (2011) *Metadaten und Standards* S. Büttner, H.-C. Hobohm, & L. Müller, eds., Bad Honnef: BOCK + HERCHEN Verlag. URL: http://opus4.kobv.de/opus4-fhpotsdam/frontdoor/index/index/docId/208 [Zuletzt abgerufen am 17. August, 2013].

Jones, R. (2006) Institutional Repositories. In K. Garnes, A. Landøy, & A. Repanovici, eds. *Aspects of the Digital Library.* Laksevåg: Alvheim & Eide, S. 111–126.

Kähler, U. (2010) Die Dienste der Föderation DFN-AAI (Anforderungen an das IdM, Attribute der Föderation, rechtliche Fragen). In *11. Shibboleth-Workshop*. Martin-Luther-Universität, Halle.

Kersting, T. & Rauschenbach, J. (2008) eduGAIN verbindet Föderationen. In *DFN-Forum Kommunikationstechnologien*. 1. DFN-Forum Kommunikationstechnologien: Verteilte Systeme im Wissenschaftsbereich, 28.05. - 29.05.2008 in Kaiserslautern. Bonn: Gesellschaft für Informatik, S. 45–52.

Kiesche, E. & Wilke, M. (2011) E-Mail und Internet am Arbeitsplatz: Kontrollen, Mitbestimmung und Datenschutz. *Computer und Arbeit* 4.

Klimpel, P. (2012) *Freies Wissen dank Creative-Commons-Lizenzen. Folgen, Risiken und Neben- wirkungen der Bedingung "nicht-kommerziell - NC,"* Berlin: Wikimedia Deutschland, iRights.info. URL: http://irights.info/userfiles/CC-NC_Leitfaden_web.pdf [Zuletzt abgerufen am 5. April, 2014].

Klimpel, P. & Keiper, J. eds. (2013) *Was bleibt? Nachhaltigkeit der Kultur in der digitalen Welt*, iRights.Media. URL: http://files.dnb.de/nestor/weitere/collab_was_bleibt.pdf [Zuletzt abgerufen am 27. September, 2013].

Klotz-Berendes, B. & Schönfelder, G. (2000) Sicherungsverfahren für den Betrieb eines Dokumentenservers. Anforderungen, kryptographische Grundlagen, Zertifizierung und digitale Signatur. In B. Tröger, ed. *Wissenschaft online: Elektronisches Publizieren in Bibliothek und Hochschule*. Frankfurt/Main: Vittorio Klostermann, S. 214–228.

Klump, J. (2010) Digitale Forschungsdaten. In H. Neuroth, A. Oßwald, R. Scheffel, S. Strathmann, & K. Huth, eds. *nestor-Handbuch: Eine kleine Enzyklopädie der digitalen Langzeitarchivierung*. Boizenburg: Werner Hülsbusch, S. 104–115.

Kretzer, S. (2013) Infrastruktur für qualitative Forschungsprimärdaten – Zum Stand des Aufbaus eines Datenmanagements von Qualiservice. In D. Huschka, H. Knoblauch, C. Oellers, & H. Solga, eds. *Forschungsinfrastrukturen für die qualitative Sozialforschung*. Berlin: Scivero-Verlag, S. 93–112.

Kruse, R. (2001) *Entwicklung eines Werkzeugs für die Administration eines Trouble Ticket Systems*, Hagen: Forschungsberichte des Fachbereichs Elektrotechnik und Informationstechnik. Fernuniversität Hagen.

Kupietz, M. & Lüngen, H. (2014) Recent Developments in DeReKo. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*. Reykjavik: European Language Resources Association (ELRA).

Lepper, U. (2007) *E-Mail und Internet am Arbeitsplatz*, Düsseldorf: Landesbeauftragter für Datenschutz und Informationsfreiheit Nordrhein-Westfalen.

Lüngen, H. & Sperberg-McQueen, C. M. (2012) A TEI P5 Document Grammar for the IDS Text Model. *Journal of the Text Encoding Initiative*. URL: http://jtei.revues.org/508 [Zuletzt abgerufen am 2. April, 2014].

Lux, T. (2005) *Intranet Engineering: Einsatzpotenziale und phasenorientierte Gestaltung eines sicheren Intranet in der Unternehmung*, Springer DE.

Müller, U. & Schirmbacher, P. (2007) Der "grüne Weg zu Open Access" in Deutschland. *Zeitschrift für Bibliothekswesen und Bibliographie* 54, S. 183–193.

Murray-Rust, P. (2008) Open Data in Science. *Serials Review* 34, S. 52–64.

National Information Standards Organization (2004) *Understanding Metadata*, Bethesda, MD: NISO. URL: http://www.niso.org/publications/press/UnderstandingMetadata.pdf.

Nehrenheim, H. (2001) IT-Unterstützung im Support – Nutzung eines Trouble-Ticket-Systems. *LDVZ-Nachrichten* 1, S. 34–38.

Neuroth, H., Oßwald, A., Scheffel, R., Strathmann, S. & Huth, K. eds. (2010) *nestor-Handbuch: Eine kleine Enzyklopädie der digitalen Langzeitarchivierung*, Boizenburg: Werner Hülsbusch.

Oksanen, V., Lindén, K. & Westerlund, H. Laundry Symbols and License Management – Practical Considerations for the Distribution of LRs based on Experiences from CLARIN. In Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10). Valletta, Malta, S. 10–13.

Oßwald, A., Scheffel, R. & Neuroth, H. (2012) Langzeitarchivierung von Forschungsdaten Einführende Überlegungen. In H. Neuroth, Strathmann, A. Oßwald, R. Scheffel, J. Klump, & J. Ludwig, eds. *Langzeitarchivierung von Forschungsdaten: Eine Bestandsaufnahme*. Boizenburg: Werner Hülsbusch.

Pampel, H., Bertelmann, R. & Hobohm, H.-C. (2010) *„Data Librarianship" – Rollen, Aufgaben, Kompetenzen*, Rat für Sozial- und WirtschaftsDaten. URL: http://econpapers.repec.org/scripts/redir.pf?u=http%3A%2F%2Fwww.ratswd.de%2Fdownload%2FRatSWD_WP_2010%2FRatSWD_WP_144.pdf;h=repec:rsw:rswwps:rswwps144.

Pfeiffenberger, H. & Klump, J. (2006) Offener Zugang zu Daten - Quantensprung in der Kooperatio. *Wissenschaftsmanagement: Zeitschrift für Innovation*, S. 12–13.

Schmidt, B. (2007) Auf dem "goldenen" Weg? Alternative Geschäftsmodelle für Open-Access-Primärpublikationen. *Zeitschrift für Bibliothekswesen und Bibliographie* 54, S. 177–182.

Schwens, U. & Liegmann, H. (2004) Langzeitarchivierung digitaler Ressourcen. In Laisiepen, E. Lutterbeck, & K.-H. Meyer-Uhlenried, eds. *Grundlagen der praktischen Information und Dokumentation*. München: Saur, S. 567–570.

Spindler, G. ed. (2006) *Rechtliche Rahmenbedingungen von Open Access-Publikationen*, Göttingen: Universitätsverlag Göttingen.

Stempfhuber, M. (2009) Die Rolle von "open access" im Rahmen des wissenschaftlichen Publizieren. In *Publikationsverhalten in unterschiedlichen wissenschaftlichen Disziplinen*. Bonn: Alexander von Humboldt-Stiftung, S. 116–131.

Uhr, W., Esswein, W. & Schoop, E. (2003) *Wirtschaftsinformatik 2003: Medien, Märkte, Mobilität*, Heidelberg: Springer.

Weston, S. & Kretschmer, M. (2012) *Open Standards in Government IT: A Review of the Evidence*, Bournemouth University: Centre for Intellectual Property Policy & Management, Bournemouth University. URL: http://www.cippm.org.uk/pdfs/cippm-open-standards-final-draft-10-september-2012.pdf [Zuletzt abgerufen am 18. April, 2014].

Wissenschaftsrat (2011a) *Empfehlungen zu Forschungsinfrastrukturen in den Geistes- und Sozialwissenschaften*, Berlin. URL: http://www.wissenschaftsrat.de/download/archiv/10465-11.pdf [Zuletzt abgerufen am 9. April, 2014].

Wissenschaftsrat (2011b) *Übergreifende Empfehlungen zu Informationsinfrastrukturen*, Berlin. URL: http://www.wissenschaftsrat.de/download/archiv/10466-11.pdf [Zuletzt abgerufen am 9. April, 2014].

Wissenschaftsrat (2012) *Empfehlungen zur Weiterentwicklung der wissenschaftlichen Informationsinfrastrukturen in Deutschland bis 2020*, Berlin. URL: http://www.wissenschaftsrat.de/download/archiv/2359-12.pdf [Zuletzt abgerufen am 9. April, 2014].